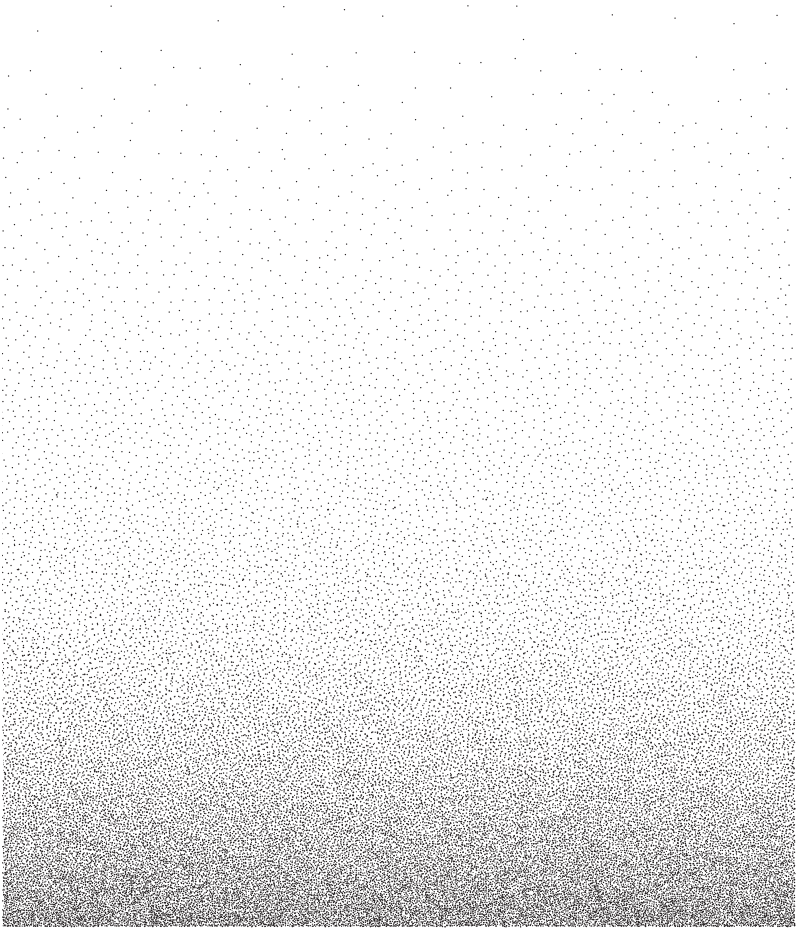
# Discovery and engineering of enzymes for chemoenzymatic peptide synthesis

Toplak, Ana

Ana Toplak

# Discovery and engineering of enzymes for chemoenzymatic peptide synthesis



Ana Toplak

# Discovery and engineering of enzymes for chemoenzymatic peptide synthesis

rijksuniversiteit
groningen

# Discovery and engineering of enzymes for chemoenzymatic peptide synthesis

Proefschrift

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. E. Sterken
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

vrijdag 25 november 2016 om 09.00 uur

door

Ana Toplak

geboren op 11 augustus 1981
te Varaždin, Kroatië

ISBN: 978-90-367-9402-2

Graphic design & layout by Remote Forms — Printed by robstolk — 2016.

The research done in this thesis was carried out at the Groningen Biomolecular Sciences and Biotechnology Institute (GBB) in the Biochemical laboratory of the University of Groningen according to the requirements of the Graduate School of Science (Faculty of Mathematics and Natural Sciences, University of Groningen) and was supported by Netherlands Organization for Scientific Research (NWO). This project is part of the Integration of Biosynthesis and Organic Synthesis program (IBOS-2; project 053.63.014) funded by the Advanced Chemical Technologies for Sustainability (ACTS) and NWO.

**Promotor**
Prof. dr. D.B. Janssen

**Beoordelingscommissie**
Prof. dr. O.P. Kuipers
Prof. dr. D.J. Slotboom
Prof. dr. R.M. J. Liskamp

## TABLE OF CONTENTS

CHAPTER — 01

# Introduction to peptide synthesis - Methods and characteristics

Ana Toplak and Dick B. Janssen

Biochemical Laboratory, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, 9747 AG Groningen, the Netherlands

## Bioactivity of peptides

Many paramount physiological and biochemical functions of life are regulated by small peptides and proteins. The range of activities attributed to peptides is extremely broad: from antibiotic and signalling activities of small peptides to catabolism, biosynthesis and regulation of cell division governed by large protein complexes. Whereas large proteins often have a role in structure, metabolism, or motion, many small peptides are involved in metabolic regulation or defense, which has stimulated the interest of the scientific community in their mode of action. Novel peptide receptors have been discovered and the bioactivity of their cognate peptides has been investigated in search for possible therapeutic applications. A wide range of peptide-based drugs are applied in the treatment of metabolic disorders and infectious diseases as well as in cancer therapy[1]. The interest in bioactive peptides is not restricted to pharmaceutical uses, but also includes applications in the food and cosmetics industries[2-5].

Currently, over 60 synthetic therapeutic peptides are present on the market[6]. Furthermore, more than 128 peptide therapeutics are in the clinical pipeline[7], and with a rate of approval that is twice that of small-molecule drugs[4], the global market for peptide therapeutics is expected to grow steadily and reach over $25 billion by 2018[8]. A number of top selling peptide therapeutics is listed in Table 1.

The development of new applications for peptide drugs requires research on new delivery systems, studies on toxicity, efficacy, distribution and metabolism, as well as clinical research, all of which are dependent on effective methods for peptide synthesis. Large doses may be required and the high costs of current production schemes limit the availability of peptide drugs. Thus, besides pharmacokinetic properties also production methods determine the possibility to develop and use peptides as drugs. The main methods for peptide synthesis are briefly reviewed below, with a focus on chemoenzymatic processes consisting of enzymatic coupling of chemically prepared or activated amino acids and peptides. A key step in chemoenzymatic peptide synthesis is the selection, discovery or engineering of suitable coupling enzymes, and therefore various strategies to obtain peptide coupling and modification enzymes will be discussed in Chapter 2.

**TABLE 1.**

Top selling peptide therapeutics (2009-2011)
and their synthesis method[6,7]

Abbreviations: AzGly = azaglycine, Thol = threoninol, O$^t$Bu = t-butyl ester, Et = ethyl group, Pyr = pyroglutamic acid

| Brand name | International nonproprietary names | Sales in 2011 (in billion US$) | Length | Sequence | Indications | Synthesis method |
|---|---|---|---|---|---|---|
| Copaxone | glatiramer acetate | 4.18 | random mixture | H-(Glu, Ala, Lys, Tyr)$_n$-OH | relapsing-remitting multiple sclerosis | chemical |
| Lupron | leuprorelin | 2.27 | 9 aa | Pyr-His-Trp-Ser-Tyr-D-Leu-Leu-Arg-Pro-NHEt | advanced prostate cancer, breast cancer | chemical |
| Sandostatin | octreotide acetate | 1.44 | 8 aa | H-D-Phe-c[Cys-Phe-D-Trp-Lys-Thr-Cys]-Thol, | acromegaly, carcinoid syndrome | chemical |
| Zoladex | goserelin acetate | 1.19 | 10 aa | Pyr-His-Trp-Ser-Tyr-D-Ser(O$^t$Bu)-Leu-Arg-Pro-AzGly-$NH_2$ | advanced prostate cancer, breast cancer | chemical |
| Victoza | liraglutide | 1.11 | 31 aa | H-His-Ala-Glu-Gly-Thr-Phe-Thr-Ser-Asp-Val- Ser-Ser-Tyr-Leu-Glu-Gly-Gln-Ala-Ala-N$^{\varepsilon}$-[N- (1-oxo-hexadecyl)-L-γ-Glu]-Lys-Glu-Phe-Ile-Ala-Trp-Leu-Val-Arg-Gly-Arg-Gly-OH | diabetes mellitus type 2 | chemical/ recombinant |
| Forteo | teriparatide | 0.95 | 35 aa | H-Ser-Val-Ser-Glu-Ile-Gln-Leu-Met-His-Asn-Leu-Gly-Lys-His-Leu-Asn-Ser-Met-Glu-Arg-Val-Glu-Trp-Leu-Arg-Lys-Lys-Leu-Gln-Asp-Val-His-Asn-Phe-OH | osteoporosis | recombinant |
| Byetta | exenatide | 0.68 | 39 aa | H-His-Gly-Glu-Gly-Thr-Phe-Thr-Ser-Asp-Leu- Ser-Lys-Gln-Met-Glu-Glu-Glu-Ala-Val-Arg-Leu-Phe-Ile-Glu-Trp-Leu-Lys-Asn-Gly-Gly-Pro-Ser-Ser-Gly-Ala-Pro-Pro-Pro-Ser-$NH_2$ | diabetes mellitus type 2 | chemical |

# Peptide synthesis

## Chemical synthesis

There is no single best technology for the production of peptides. Current methods for peptide production are chemical synthesis, production in recombinant systems, and chemo-enzymatic synthesis, and the method of choice depends on sequence, length and properties of the peptide. At this moment, chemical peptide synthesis is still the preferred method for the production of therapeutic peptides, since it allows the synthesis of almost any desired peptide sequence, including peptides with non-proteinogenic amino acids. In general, chemical peptide synthesis includes selection of protecting groups, a deprotection and/or activation method, and peptide bond formation step[9]. Chemical peptide synthesis and coupling reactions are highly sequence dependent, and both method development and analysis become more critical as peptides grow in complexity[4]. Current peptide-based products are at the smaller end (8–10 amino acids) of the size spectrum of bioactive peptides, but peptides up to 45 amino acids are also on the market[10]. Synthetic routes can become problematic due to low segment solubility and also racemization might be a problem. The number and amount of impurities, many of which are very similar to the desired product, increase with the number of amino acids, which dramatically influences production costs[4]. Advantages and limitations of the available methods are discussed below (Table 2).

TABLE 2.

Methods for peptide synthesis

| | Chemical methods | | | Chemoenzymatic methods | Enzymatic methods | |
|---|---|---|---|---|---|---|
| | Classical | | | | | |
| | Solution peptide synthesis | Solid-phase peptide synthesis | Hybrid approach/ segment condensation | Chemoenzymatic peptide synthesis | Fermentative peptide synthesis | |
| | | | | | Ribosomal | NRPS |
| Scale | g-ton | mg-ton | g-ton | mg | mg | g-ton |
| Peptide length | short-medium | medium-large | medium-large | Medium-very large | medium | short-large |
| Protection required | partial-total | total | total | none | none | minimal-none |
| Sequence versatility | medium-high | high | high | narrow | narrow | medium-high |
| Non-proteinogenic amino acids | yes | yes | yes | yes | yes | yes |
| Racemization | some | none | yes | none | none | none |
| Reaction medium | organic solvent | organic solvent | organic solvent | aqueous medium | aqueous medium | organic solvent, organic solvent-water mixtures, aqueous medium |
| R&D effort | small | very small | large | large | large | large |
| Development phase | mature | mature | mature | embryonic | embryonic | infantry |
| Production costs | high | very high | high -very high | medium | medium-high | medium |
| Environmental impact | high | very high | high | some | some | some |
| Industrial scale | yes | yes | yes | no | no | yes |

## Solution-phase peptide synthesis

Solution phase synthesis (SPS) was the first method developed for peptide synthesis, and can include both stepwise coupling reactions and coupling of peptide segments. Different versatile coupling methods are available[11], with the carbodiimide method being commonly used for peptide segment coupling[12]. Here, the coupling reagent *N',N''*-dicyclohexylcarbodiimide (DCC) activates the carboxylate function of an $N^{\alpha}$-protected amino acid that once activated is able to form a peptide bond with an amine nucleophile. A coupling additive, e.g. 1-hydroxybenzotriazole (HOBt) is added to prevent racemization and improve efficiency. Solution phase peptide synthesis is a scalable method that is mainly applied for the synthesis of small peptides (<10 amino acids)[13,12]. Drawbacks are the need to isolate and purify the intermediates after each step and the requirement for protecting groups. Especially for longer peptides (> 10 amino acids) this renders the synthesis costly and time-consuming. Another problem is that racemization of the C-terminal amino acid in solution may occur, and preferred coupling positions therefore are Gly-Xxx or Pro-Xxx[14]. In addition, as the size of the target peptide increases, the solubility of the segments to couple becomes a major issue[15].

## Solid-phase peptide synthesis

The pioneering work of Merrifield[16], who introduced solid-phase peptide synthesis (SPPS), revolutionized the field of chemical peptide synthesis, especially because it allowed the development of automated synthetic processes. Today, SPPS is the method of choice for the synthesis of small- to medium-size peptides (<50 amino acids). The desired sequence is assembled in a linear fashion from the C-terminus to the N-terminus (C→N direction) by repetitive cycles of $N^{\alpha}$-de-protection and coupling steps (Figure 1). Peptide coupling reagents and coupling additives are required for carboxylic acid activation and peptide bond formation. Solid-phase synthesis is fast and generally racemization free because the coupling occurs between C-terminus of an activated amino acid and the N-terminus of the peptide chain attached to the solid support.





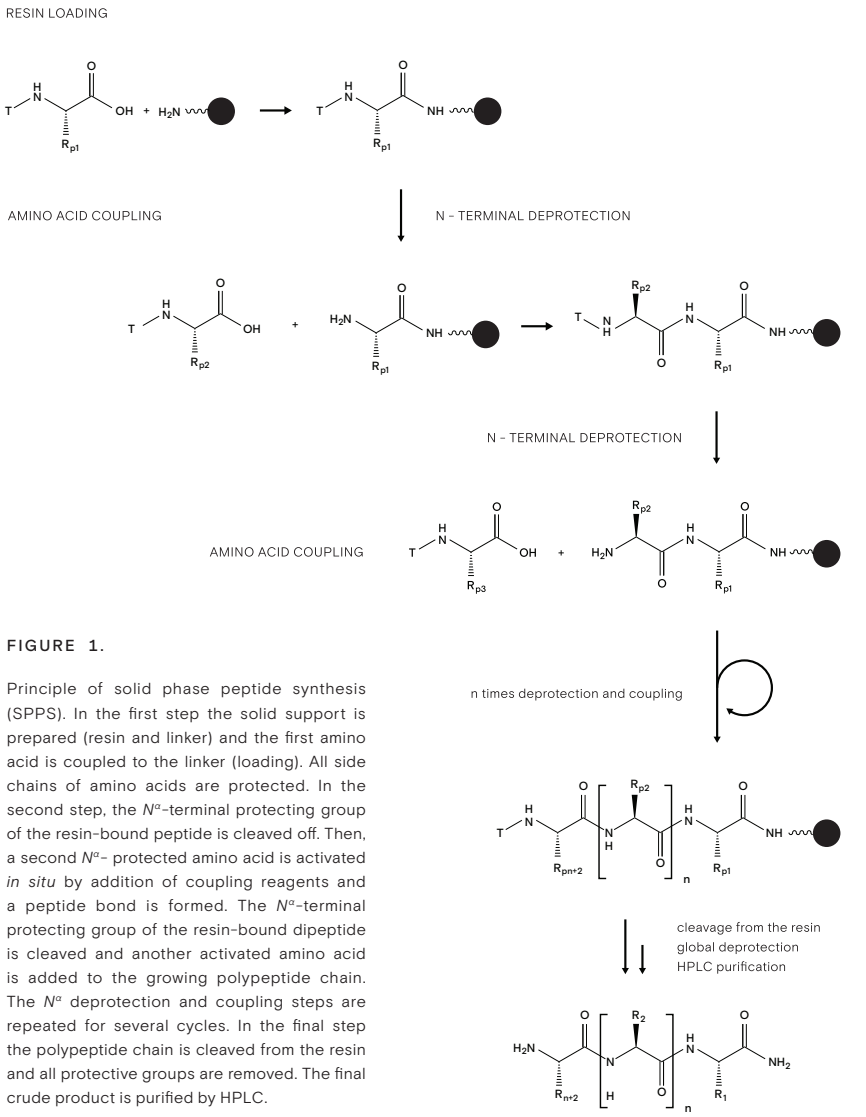



**FIGURE 1.**

Principle of solid phase peptide synthesis (SPPS). In the first step the solid support is prepared (resin and linker) and the first amino acid is coupled to the linker (loading). All side chains of amino acids are protected. In the second step, the $N^{\alpha}$-terminal protecting group of the resin-bound peptide is cleaved off. Then, a second $N^{\alpha}$- protected amino acid is activated *in situ* by addition of coupling reagents and a peptide bond is formed. The $N^{\alpha}$-terminal protecting group of the resin-bound dipeptide is cleaved and another activated amino acid is added to the growing polypeptide chain. The $N^{\alpha}$ deprotection and coupling steps are repeated for several cycles. In the final step the polypeptide chain is cleaved from the resin and all protective groups are removed. The final crude product is purified by HPLC.

Protective groups: T = temporary $N^{\alpha}$-terminal protective group ($^{t}$Boc = tert-butoxycarbonyl), Fmoc = 9-fluorenylmethyloxycarbonyl; R = side chain; p = permanent side chain protecting group (Bzl = benzyl, $^{t}$Bu = tert-butoxy); A = activating group; n = number of cycles.

Common strategies involve combinations of temporary (T) and permanent ($P_n$) protecting groups e.g. $^t$Boc/Bzl (T/$P_n$) and Fmoc/$^t$Bu (T/$P_n$). The latter is nowadays widely used, even on industrial scale, because temporary and permanent protecting groups are removed by different mechanisms (i.e. Fmoc is base labile and side chain protective groups are acid labile) which makes it possible to use milder conditions for removal of $N^{\alpha}$–terminal protecting group, final cleavage from the resin and side-chain deprotection than in the case where protective groups are not orthogonal. A summary of general strategies and protocols for solid phase peptide synthesis is described in a review by Amblared et al.[17]

Disadvantages of solid-phase peptide synthesis are that in order to reach high yields per cycle, reagents need to be used in excess and washed away, resulting in production of large amounts of reagents and solvent as waste (e.g. 1000 L solvent/kg of peptide drug[12]). Tedious and expensive preparative HPLC purification is needed to remove impurities with similar properties, which further increases production costs. In addition, peptides longer than 10 amino acids tend to form tertiary structures due to a process called hydrophobic collapse, making the further synthesis difficult.

## Chemical solution-phase peptide segment condensation

In the so-called hybrid approach, solution-phase reactions are used to couple peptide segments that are usually obtained by step-wise solid-phase synthesis (Figure 2). This way, the advantages of the solid-phase and solution-phase methods are combined, and it is often an attractive way to synthesize longer peptides[18]. The selection of coupling sites is strongly sequence dependent. Ideally, the starting peptide segments (i.e. 10-mers) are obtained by means of solid-phase peptide synthesis, and are then coupled in convergent manner (e.g. 10+10+10) using solution-phase chemistry. To avoid racemization Gly or Pro should be the C-terminal amino acids of the peptide segments. If that is not the case, the purification becomes troublesome and expensive. Thus, critical issues are solubility of the intermediates, undesired epimerization at the coupling positons, and troublesome isolations.

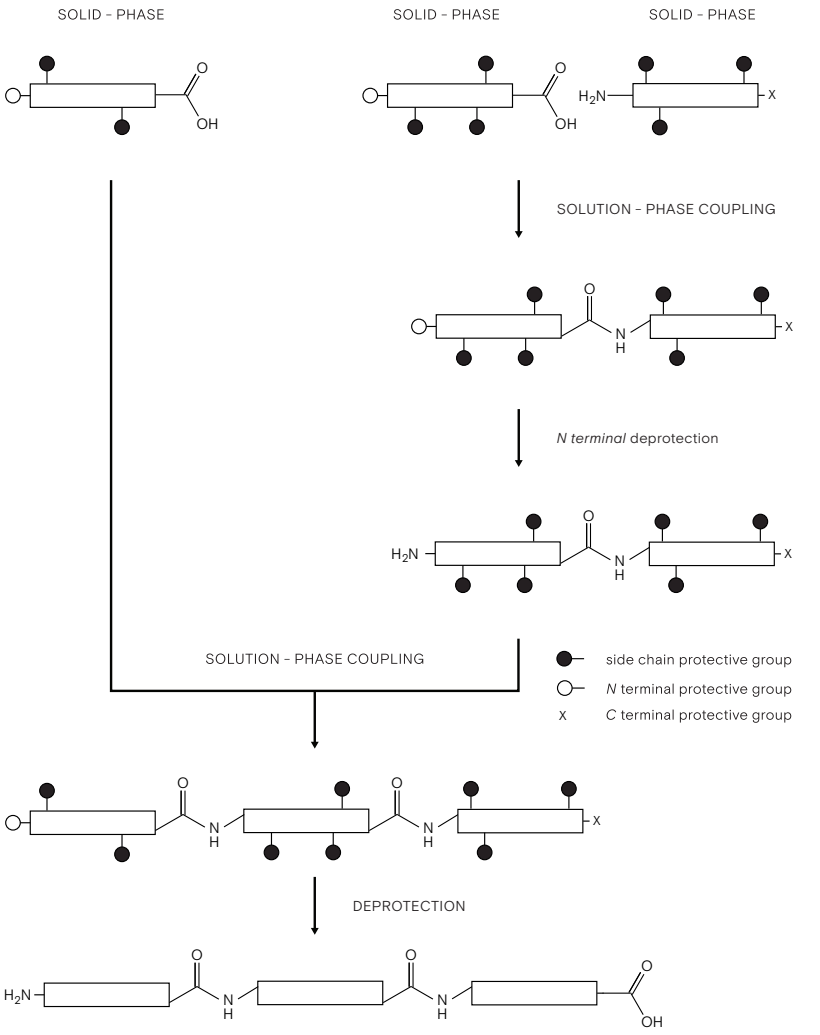



**FIGURE 2.**

Hybrid approach for the synthesis of a 30-mer peptide in *C*- to *N*- direction. Symmetric peptide segments (decapeptides) obtained by solid-phase peptide synthesis are coupled in a stepwise manner (10+10+10), using solution-phase chemistry. The selection of sites for the coupling reactions is crucial for process efficacy in terms of minimizing the risk of racemization and purification costs.

## Native chemical ligation

Native chemical ligation (NCL) is a relatively new synthetic method that enables covalent condensations of large unprotected peptide segments that are soluble in buffer systems. The method exploits the unique properties of a thioester group. The coupling of an activated acyl donor peptide in a form of a thioester and a nucleophile (peptide with N-terminal cysteine) proceeds in two steps. The first step involves a reversible thiol-thioester exchange: the side-chain SH of the N-terminal amino acid (usually a cysteine) of the amino donor peptide attacks the activated thioester of the acyl donor. This is followed by spontaneous irreversible S → N acyl transfer to the final product that has a native peptide bond at the ligation site (Figure 3). NCL has been used to make synthetic proteins up to 200 amino acids long[19].

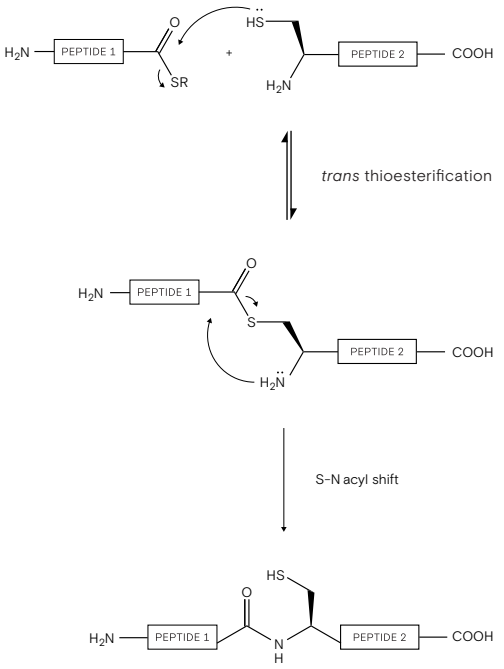



FIGURE 3.

Principle of native chemical ligation: thioester-mediated amide-forming chemoselective ligation of two unprotected peptide segments. Under denaturing conditions at neutral pH and in the presence of an exogenous thiol catalyst, the first step (thiol-mediated thioester exchange) is reversible, while the second step (S-to-N acyl shift) is irreversible and gives a native peptide bond at the ligation site.





The requirement for a cysteine at a coupling site is often regarded as the main limitation of NCL. In addition, the need for an expensive thioester segment on the acyl donor limits the large-scale applicability of this method. Even though modified methods were developed to circumvent the need for a cysteine at the coupling site (desulfurization, auxiliary group), NCL is not applied on industrial scale[20]. Novel methods that address the thioester limitation have been developed, such as expressed protein ligation (EPL), which allows α-thioester segments to be produced recombinantly as engineered intein-domain fusion peptides that can later be cleaved by thiolysis and coupled to a desired target using NCL[21,22] (Figure 4). However, the method depends on N-terminal cysteine residues and has restrictions similar to NCL. Besides recombinant production of α-thioester segments, where only natural amino acids can be incorporated, α-thioester peptides with desired modifications can be chemically prepared using peptide hydrazides that can be converted to thiosters[23] or by applying enzymatic C-terminal modification to thioester[24,25].

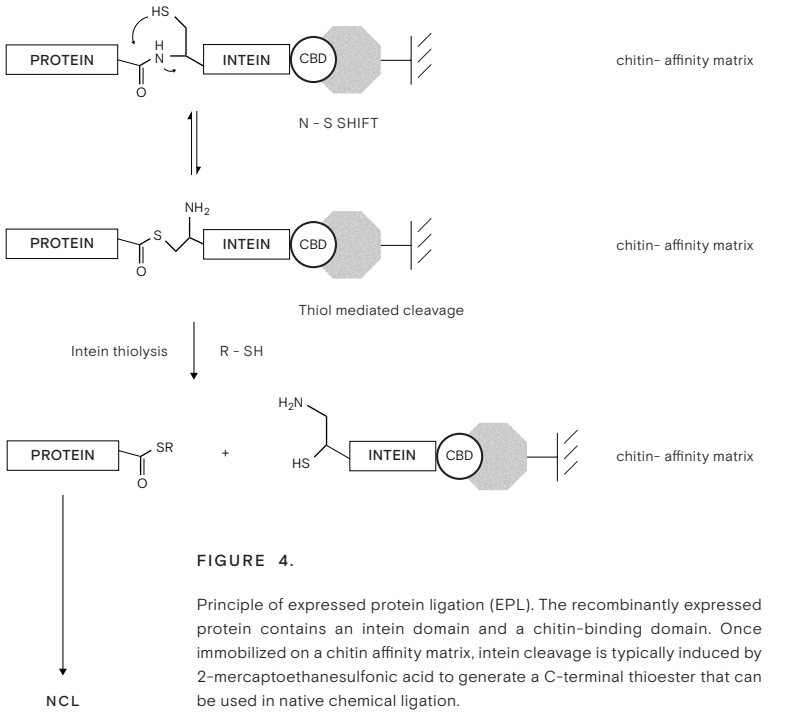



FIGURE 4.

Principle of expressed protein ligation (EPL). The recombinantly expressed protein contains an intein domain and a chitin-binding domain. Once immobilized on a chitin affinity matrix, intein cleavage is typically induced by 2-mercaptoethanesulfonic acid to generate a C-terminal thioester that can be used in native chemical ligation.

In expressed enzymatic ligation (EEL) the EPL strategy is combined with a substrate-mimetic strategy. Peptide segments are produced by recombinant expression as in EPL and are coupled by a protease using a substrate mimetic approach[26] (see Chapter 2). This ligation approach eliminates the need for cysteines at the ligation site, but as with all proteases care has to be taken to prevent possible hydrolysis of the peptide segments or the final product.

## Fermentative peptide synthesis

Biochemical peptide synthesis employing the cell's natural capacity to synthesize peptides is attractive because of high chemo- and regioselectivity, lack of racemization, mild process conditions and low environmental impact[27,28]. Two methods can be distinguished: ribosomal peptide synthesis, based on transcription and translation of DNA coding sequences, and non-ribosomal peptide synthesis, employing large modular protein complexes. Unfortunately, the standard application of ribosomal peptide synthesis by genetically engineered cells is restricted to proteinogenic amino acids and to relatively long peptides (>50 amino acids[29]) because short peptides are easily degraded in the metabolic pathways of the recombinant host. Strategies for fermentative production of shorter peptides and peptides containing non-natural amino acids are being developed as well[2,30]. The restriction that fermentation is feasible only in case of relatively long peptides composed of natural amino acids and the high development costs limit the widespread use of ribosomal peptide synthesis. As a result, ribosomal peptide synthesis is only applied in the late drug development stage and is used only for the synthesis of large peptides such as insulin[31] and human growth factor[32].

To surpass the limitations of the host, cell-free translation systems can be considered. The system includes a DNA template that encodes the desired peptide, a ribosomal extract as the synthesis machinery, amino acids, and an energy source. Although such systems better tolerate non-natural amino acids, the low product yield, low RNA template stability and low translation efficiency require significant improvement before large-scale application is possible[33].

Whole-cell non-ribosomal peptide synthesis is mediated by dedicated multimeric enzyme complexes that direct the synthesis of versatile bioactive compounds such as antifungal and antibiotic molecules, independent of the ribosomes.





These non-ribosomal peptide synthetases (NRPS) are of particular interest since they produce bioactive polypeptides of various size and shape, including peptides carrying modifications such as N-methylation, cyclization, and the presence of D-amino acids. NRPSs are employed for large scale peptide synthesis, such as β-lactam antibiotic and cyclosporin production. However, they are not yet applied for the synthesis of non-native peptides, which is mainly due to the inability to tailor and re-design NRPSs for synthesis of a desired peptide sequence[34,35].

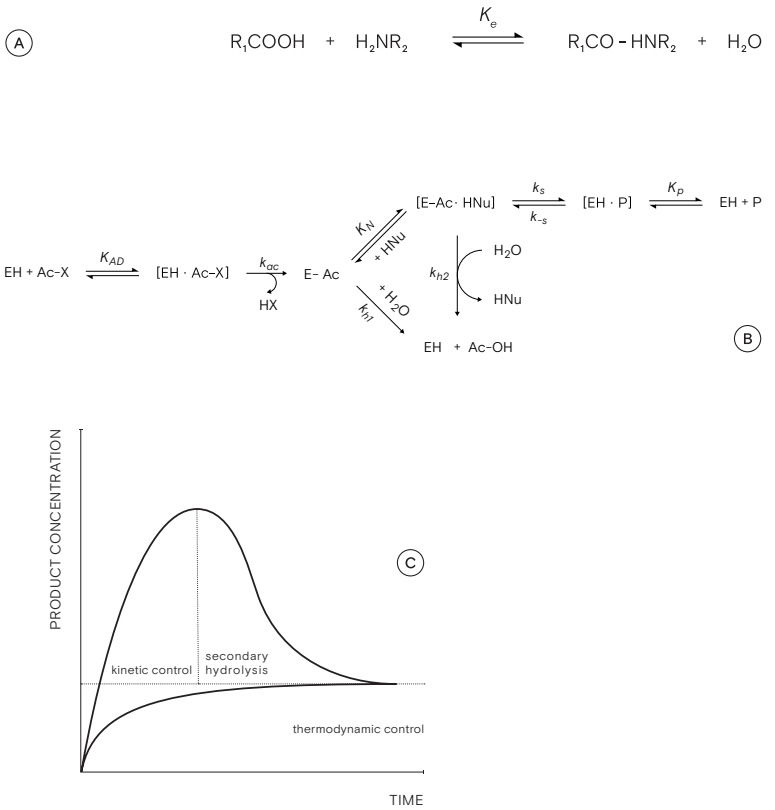
## Enzymatic and chemoenzymatic peptide synthesis

### Peptide bond synthesis by proteases

Enzymatic peptide synthesis is based on the application of proteases or related hydrolytic enzymes in the reverse direction. The possibility to use proteases for peptide synthesis under specific reaction conditions was first predicted by Van't Hoff[36] and Ostwald[37] in 1898. and demonstrated by Bergmann and Fraenkel-Conrat some 40 years later[38]. Both stepwise coupling and segment ligation can be catalyzed by proteases. In the 1980s, the use of hydrolases in the reverse direction became a field of intensive research, mostly due to the work of Klibanov and coworkers who used organic solvents to trigger coupling reactions[39,40]. Since then, and stimulated by the growing importance of bioactive peptides, a lot of research has been devoted to finding better enzymes and optimal conditions for peptide synthesis. Modern approaches such as genome mining and protein engineering have been used for discovery and engineering of peptidases that can be used for synthesis.

Peptidases can catalyze the synthesis of a peptide bond in either thermodynamically or kinetically controlled manner[41,42]. The thermodynamic approach is a reversal of hydrolysis which is achieved by modifying reaction conditions. This approach is useful only if a shift of the chemical equilibrium occurs, e.g. by reducing water activity (solvent-free systems, addition of cosolvents, use of neat organic solvent) or by removing the product (extraction into organic layer, separation by molecular traps, precipitation)[43]. If peptides or amino acids with free carboxylate and amine groups are coupled, solvent hydrophobicity will stimulate synthesis, and the reaction will proceed to a degree of conversion that is dependent on the conditions used, especially solvent composition and pH (Figure 5).

Kinetically controlled peptide synthesis is based on the capacity of some proteases to form an acyl-enzyme intermediate that can undergo, under appropriate conditions, deacylation by the N-terminus of an amino acid or peptide rather than cleavage by water. Suitable enzymes for kinetically controlled coupling are serine or cysteine proteases. The acyl donor is added in the form of an activated precursor that forms the acyl-enzyme intermediate and provides the N-terminal segment of the product peptide. The amine group of a nucleophilic peptide or amino acid reacts in an aminolytic reaction with the acyl enzyme and becomes the C-terminal segment (Figure 5).








In a kinetically controlled process, the nucleophile needs to act in competition with water to form a peptide bond instead of hydrolysis of the acyl-enzyme. Different nucleophiles can compete with water, e.g. an alcohol yields an ester in a trans-esterification reaction; a peroxide as nucleophile generates a peracid, and an amine nucleophile yields an amide. Enzymatic formation of hydrazides is also possible, as well as the formation of derivatives of hydroxamic acid (Figure 6). Compared to the thermodynamic approach, the kinetically controlled synthesis enables higher conversions and requires less enzyme, provided a good enzyme is available[27]. In principle, product yield over time goes through a maximum, and since product hydrolysis can also happen, the reaction will slowly proceed to equilibrium.

Due to the competing hydrolytic reaction and the possibility of product hydrolysis, the level of (transient) product accumulation strongly depends on the enzyme properties. The enzyme should have a low tendency to hydrolyze the acyl donor, and in case of peptide coupling, also refrain from hydrolyzing bonds in the peptide building blocks. An important parameter is the aminolysis to hydrolysis ratio (or synthesis-hydrolysis ratio, S/H), which should be as high as possible. In addition, the ideal coupling enzyme is robust in terms of its operational stability, including temperature stability and co-solvent tolerance.

**FIGURE 5.**

Principle of enzymatic peptide synthesis. **A** Synthesis under thermodynamic control is based on microscopic reversibility. The overall equilibrium constant $K_e$ will be strongly dependent on ionization constants of the different species and the reaction pH. **B** Synthesis under kinetic control. The activated acyl donor (Ac-X) forms a Michaelis complex [EH•Ac-X] that reacts to the acyl-enzyme intermediate E-Ac. The deacylation of this intermediate can be performed either by water yielding hydrolytic product Ac-OH and free enzyme, or by an added nucleophile (HNu) that leads to the synthetic product P and free enzyme via a second Michaelis complex [E-Ac•HNu]. This complex may still hydrolyze if water competes with bound nucleophile. Synthetic yields are determined by rate constants $k_{ac}$, $k_{h1}$, $k_{h2}$, $k_s$, $k_{-s}$ and binding constants $K_{AD}$, $K_N$ and $K_p$. **C** Schematic drawing of the time course of formation of synthetic product.

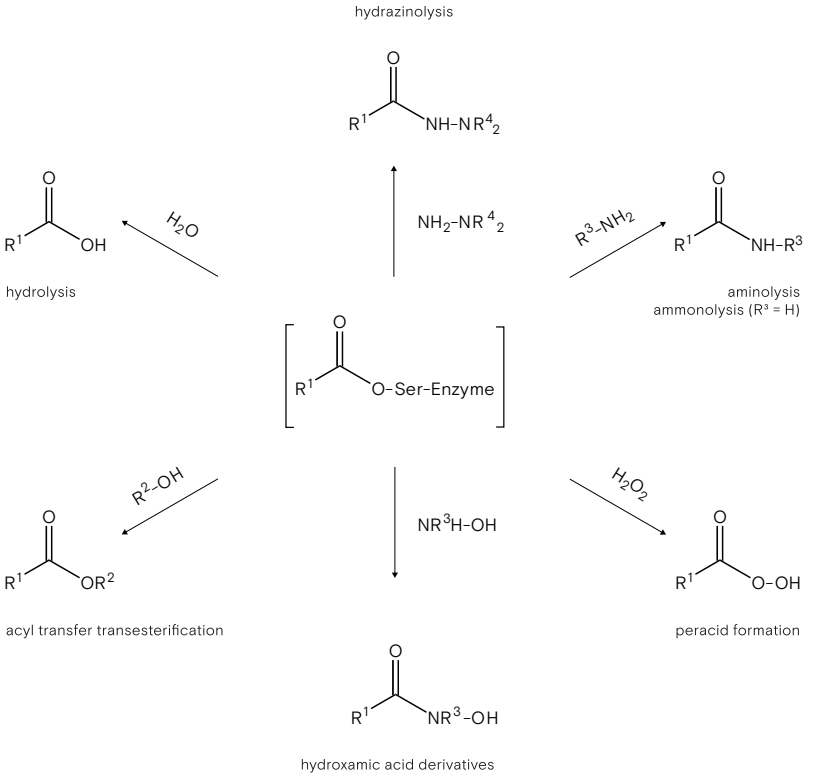



FIGURE 6.

Biocatalytic applications of serine peptidases. In a low water activity environment a variety of nucleophiles may outcompete water for cleaving the acyl-enzyme intermediate allowing formation of various synthetic products. Transesterification may occur with alcohol as a nucleophile, whereas an amine nucleophile leads to the formation of an amide bond. Other nucleophiles may yield hydroxamic acids derivatives or hydrazides. Hydrogen peroxide as nucleophile leads to peracid formation.





An important enzyme property is also substrate specificity. Restricted substrate specificity, such as encountered in trypsin and chymotrypsin, can be an advantage since it reduces the number of possibilities for hydrolysis of a synthetic product, but also a disadvantage, since it reduces the flexibility in terms of peptide bonds that can be synthesized. Peptidases can be distinguished on basis of their positional specificity and recognition sequence. The precise cleavage site is mainly determined by favorable interactions between side chains of the amino acids of the peptide (designated P2, P1, P1', P2' etc.) and specific subsites in the active site (S2, S1, S1', S2' etc.) of the protease. The notation of sites follows the Schechter and Berger nomenclature[44], according to the scheme shown in Figure 7[45].

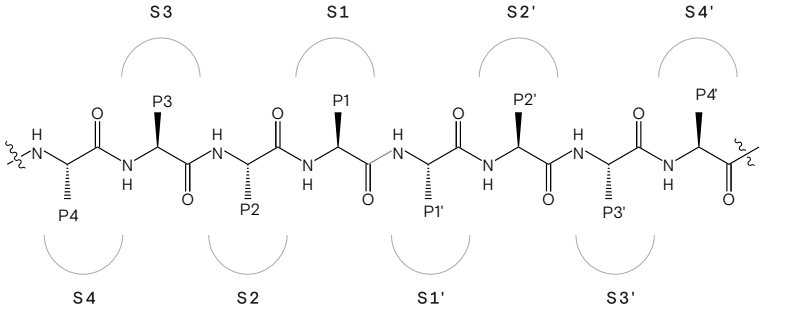



FIGURE 7.

Schechter and Berger nomenclature for the active site of a peptidase and its substrate. The scissile bond is located between P1 and P1'.

Although fully enzymatic step-wise synthesis of peptides is possible, in practice a convergent approach consisting of a combination of chemical and enzymatic steps is preferred. This requires the target sequence to be divided into segments that are independently produced (usually chemically) and can be coupled enzymatically. This causes a lack of general applicability, and in combination with the necessity to use expensive activated precursors for kinetic coupling caused reluctance to use this chemo-enzymatic peptide synthesis strategy on industrial scale.

Enzymatic synthesis of peptides by a sequential approach (stepwise addition of a single amino acid to the growing peptide chain), or by an economically more favorable convergent approach where larger peptide segments are enzymatically coupled, has been reported, both in water as well as in organic solvent mixtures, applying kinetically or thermodynamically controlled synthesis[46,47]. For example, the pentapeptide precursor (Cbz-Tyr-Gly-Phe-Gly-Gly-OEt) of the bioactive osteogenic growth protein was synthesized in 50% yield (3+2 segment coupling with papain or chymotrypsin), or in 70% yield (chymotrypsin- and thermolysin-mediated coupling in almost dry organic solvent)[48].

## Reaction medium design

Since the work of Klibanov and colleagues on the use of enzymes in organic solvents, the field of non-aqueous enzymology has developed rapidly[39,40]. The possibility to apply enzymes in water-cosolvent systems, including two-phase systems, and the option to use neat organic solvents opened a whole range of synthetic applications of enzymes that naturally hydrolyze peptide bonds. The presence of cosolvents or the use of neat solvents may have a profound effect on reaction equilibria, making it possible to use hydrolytic enzymes for coupling reactions. Furthermore, the use of organic solvents may have advantages such as high substrate- and product solubility and the partitioning of substrates and products over different phases. On the other hand, organic solvents may cause molecular or (inter)phase toxicity, and disturb the stable conformation of enzymes.

The use of cosolvents to increase peptide substrate solubility has been described both for thermodynamic coupling reactions and for kinetically controlled synthesis[49,50]. Moderate amounts of water-miscible cosolvents were to improve substrate solubility, without considerable negative effect on the enzyme[51]. The solubility of peptides in a solvent or buffer is influenced by their polarity, which is determined by the amino acid composition and the conformation of the peptide. Accordingly, peptide solubility can be enhanced by selecting appropriate solvent mixtures. For example, addition of the strongly solvating compound dimethylformamide (DMF) or dimethylsulfoxide (DMSO) is known to increase solubility of peptides[52].





Anhydrous organic solvents are a promising medium for peptide synthesis under kinetic control, since hydrolysis is avoided. Furthermore, by using a suitable solvent the selectivity and activity of a peptidase can be modulated. Only a very small amount of water is necessary for an enzyme to have conformational flexibility and to be catalytically active, and this essential water can range from a few % in organic solvent to as little as a monolayer[53]. Moreover, the effect of organic solvent on enzyme structure is not as detrimental as previously thought, as concluded from studies on subtilisin crystals in organic solvents[54]. Hydrophobic organic solvents often increase thermal stability of enzymes. For example, the half-life of α-chymotrypsin in aqueous solution at 55°C is 15 min[55], whereas in octane at 100°C it is about 80 min[39]. Enzyme specificity, stereoselectivity, regiospecificity, and chemoselectivity are also affected[51].

The relaxed selectivity of enzymes in organic medium enables new synthetic applications of peptidases. As an example, a dipeptide containing either D-L, L-D or D-D peptide bonds can be synthesized using subtilisin in anhydrous solvent, whereas these bonds are usually not sensitive to peptidase-mediated hydrolysis[56]. This also shows that the specificity an enzyme exhibits in water is not necessarily maintained when it is used in anhydrous medium. However, enzyme activity in neat organic solvents can be 5,000-100,000 less than in water and to alleviate this loss optimized reaction conditions have to be applied (i.e. medium selection, water activity control, optimal enzyme preparation)[57].

A remarkably solvent-stable protease is subtilisin. It showed a unique tolerance to hydrophilic solvents that dissolve long protected peptides, such as tetrahydrofuran (THF), acetonitrile, or alcohols. The synthetic potential of subtilisin Carlsberg (Alcalase) was recently explored by Quaedflieg and coworkers[58–62]. In these experiments, Alcalase was used in neat organic solvent in synthetic reactions containing activated precursors for peptide synthesis, catalyzing condensation of a 10-mer and a 9-mer. Although medium engineering is primarily focused on influencing the properties of an enzyme by adding cosolvents (activity, specificity, stability), it can also involve the use of frozen aqueous solutions[63], use of ionic liquids[64] or supercritical fluids and[65–67] combinations thereof. In a recent attempt to make peptide synthesis more green, the amounts of solvent were reduced to obtain a slurry (in SPPS), or solvent was even completely omitted (ball-milling technology[68]).

## Reaction medium design

The rate and yield of peptide synthesis reactions not only depends on enzyme and reaction conditions, but also on the type of reactants. Important aspects are the choice of the activating groups in kinetically controlled synthesis, the choice of terminal protecting groups in coupling reactions, modifications of the peptide chain that influence substrate and product solubility, and the selection of ligation sites if peptides are to be coupled.

Compounds with good leaving groups such as carboxamidomethyl (Cam, glycolamide)[69] or 2,2,2-trifluoroethyl (Tfe) esters[70] proved to be better acyl donors for peptide synthesis than (m)ethyl esters (Table 3). Their enhanced activity is explained by strong electron-withdrawing properties of the leaving groups which make the carbonyl group more susceptible to nucleophilic attack by an active site serine. In addition, it is believed that some activating groups, such as in Cam esters, have the ability to bind to the enzyme via a hydrogen bond in the same fashion as an amide of a peptide backbone binds. Recently, substituted phenyl esters and carboxamidomethyl groups that are elongated with an amino acid amide were tested. These derivatives form additional binding interactions with the enzyme and gave better synthesis as compared to other activating groups[71]. Carboxamidomethyl groups and modified carboxamidomethyl activating groups are of particular interest for chemoenzymatic peptide synthesis since they give very good activity, are not prone to racemization and can be easily synthetized either chemically or enzymatically[71,72].

Recognizing the importance of binding of the acyl donor for synthetic performance in kinetically controlled coupling, modifications of the leaving group such that it better matches enzyme selectivity has been proposed[73,74] (Table 3). In the so-called inverse substrate or substrate-mimetic approach, a reactive group such as in a *p*-guanidinophenyl (OGp) ester forces formation of an acyl-enzyme intermediate even in the absence of an amino acid at the P1 position that matches the S1 pocket specificity. Instead, the S1 pocket is used to bind the leaving group, which is designed to mimic the side chain that fits in the S1 subsite, and a reactive acyl-enzyme is formed. Prior to deacetylation step the substrate changes orientation liberating S' subsites, thus allowing nucleophilic attack.





This approach enabled the acceptance of amino acids and peptide sequences that are normally not recognized by chymotrypsin or trypsin[75,76]. Work with the Glu-specific endopeptidases (V8 from *Staphylococcus aureus* and BL-GSE from *Bacillus licheniformis*) showed the use of the carboxymethyl thioester group (2-mercaptoacetate thioester) as substrate mimetic for a protease that is not active with arginine at the P1 position[77,78].

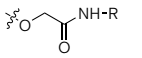
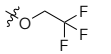
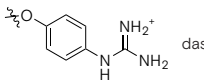
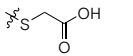
| Type | Activating group | Abbreviation | Structure | Related enzymes |
|---|---|---|---|---|
| Activating ester | carbamoylmethyl ester | Cam ester |  R = H, Cam R = amino acid, Cam derivative | serine peptidases, papain, subtiligase |
| | 2,2,2-trifluoroethyl ester | Tfe ester |  | serine peptidases |
| Substrate mimetic | *p*-guanidino-phenyl ester | OGp ester |  | arginine-specific peptidases (trypsin, chymotrypsin and papain) |
| | carboxymethyl thioesters | SCm ester |  | non-arginine-specific peptidases |

**TABLE 3.**

Strongly activating groups for kinetic peptide coupling.

### Enzyme selection and formulation

Synthetic yields and synthesis rates in enzymatic peptide synthesis can be improved by optimizing reaction conditions, by reactant optimization, and by selecting the most appropriate biocatalyst[47,70,79–85]. Especially kinetically-controlled coupling reactions are critically dependent on enzyme properties. Suitable enzyme can be acquired commercially, or they may be obtained by screening for new enzymes for nature's biodiversity. Variants of known enzymes with improved properties may be constructed by protein engineering (this thesis).

Enzyme formulation is also important, including physical or chemical modification. Especially immobilization is often applied as it contributes to enzyme stability and allows rapid separation of enzyme and reaction products, with the possibility to reuse the enzyme[86]. Due to the variations in selectivity of most proteases and differences in compatibility with reaction conditions, there is not a single peptidase that can be used for any coupling reaction. Rather, a toolbox of peptidases is needed in order to fully exploit the advantages of chemo-enzymatic peptide synthesis.





## References


01 — Thundimadathil, J. Cancer treatment using peptides: current therapies and future prospects. *J. Amino Acids* **2012**, 967347 (2012).

02 — Thayer, A. Improving peptides. *Chem. Eng. News* **89**, 13–20 (2011).

03 — Thayer, A. M. Making peptides at large scale. *Chem. Eng. News* **89**, 9–12 (2011).

04 — Lax, R. The future of peptide development in the pharmaceutical industry. *PharManufacturing Int. Pept. Rev.* 10–15 (2010).

05 — Mine, Y., Li-Chan, E. & Jiang, B. *Bioactive proteins and peptides as functional foods and nutraceuticals*. (Wiley-Blackwell, 2010).

06 — Vlieghe, P., Lisowski, V., Martinez, J. & Khrestchatisky, M. Synthetic therapeutic peptides: science and market. *Drug Discov. Today* **15**, 40–56 (2010).

07 — Kaspar, A. a & Reichert, J. M. Future directions for peptide therapeutics development. *Drug Discov. Today* **18**, 807–17 (2013).

08 — Research&Markets. *Global Peptide Therapeutics Market & Pipeline Insight*. (2014).

09 — Okada, Y. Synthesis of peptides by solution methods. *Curr. Org. Chem.* **5**, 1–43 (2001).

10 — Craik, D. J., Fairlie, D. P., Liras, S. & Price, D. The future of peptide-based drugs. *Chem. Biol. Drug Des.* 81, 136–47 (2013).

11 — Marder, O. & Albericio, F. Industrial application of coupling reagents in peptides. *ChemInform* **35**, (2004).

12 — Andersson, L., Blomberg, L., Flegel, M., Lepsa, L., Nilsson, B. & Verlander, M. Large-scale synthesis of peptides. *Biopolymers* **55**, 227–250 (2000).

13 — Guzman, F., Barberis, S. & Illanes, A. Peptide synthesis: chemical or enzymatic. *Electron. J. Biotechnol.* **10**, 279–314 (2007).

14 — Bodanszky, M. *Principles of peptide synthesis*. (Springer Berlin Heidelberg, 1993).

15 — Lloyd-williams, P. & Giralt, E. Convergeny solid-phase peptide synthesis. *Tetrahedron* **49**, 11065–11133 (1993).

16 — Merrifield, R. B. Solid phase peptide pynthesis.The synthesis of a tetrapeptide. *J. Am. Chem. Soc.* **85**, 2149–2154 (1963).

17 — Amblard, M., Fehrentz, J., Martinez, J. & Subra, G. Methods and protocols of modern solid phase peptide synthesis. *Mol. Biotechnol.* **33**, 239–254 (2006).

18 — Bray, B. L. Large-scale manufacture of peptide therapeutics by chemical synthesis. *Nat. Rev. drug Discov.* **2**, 5–9 (2003).

19 — Dawson, P. E., Muir, T. W., Clark-Lewis, I. & Kent, S. B. Synthesis of proteins by native chemical ligation. *Science* **266**, 776–779 (1994).

20 — Kent, S. B. H. Total chemical synthesis of proteins. *Chem. Soc. Rev.* **38**, 338–51 (2009).

21 — Muir, T. W., Sondhi, D. & Cole, P. A. Expressed protein ligation : A general method for protein engineering. *Proc. Natl. Acad. Sci.* **95**, 6705–6710 (1998).

22 — Machova, Z. & Beck-Sickinger, A. G. Expressed protein ligation for protein semisynthesis and engineering. *Methods Mol. Biol.* **298**, 105–130 (2005).

23 — Zheng, J.-S., Tang, S., Qi, Y.-K., Wang, Z.-P. & Liu, L. Chemical synthesis of proteins using peptide hydrazides as thioester surrogates. *Nat. Protoc.* **8**, 2483–95 (2013).

24 — Quaedflieg, P. J. L. M. & Merkx, N. S. M. Chemo-enzymatic synthesis of a C-terminal thioester of an amino acid or a peptide. Patent WO2009047354 A1 filed 12 Oct. 2008, and issued 16 Apr. 2009.

25 — Tan, X.-H., Yang, R., Wirjo, A. & Liu, C.-F. Subtiligase as a hydrothiolase for the synthesis of peptide thioacids. *Tetrahedron Lett.* **49**, 2891–2894 (2008).

26 — Machova, Z., von Eggelkraut-Gottanka, R., Wehofsky, N., Bordusa, F. & Beck-Sickinger, A. G. Expressed enzymatic ligation for the semisynthesis of chemically modified proteins. *Angew. Chem. Int. Ed. Engl.* **42**, 4916–8 (2003).

27 — Schellenberger, V. & Jakubke, H.-D. Protease-catalyzed kinetically controlled peptide synthesis. *Angew. Chemie Int. Ed. English* **30**, 1437–1449 (1991).

28 — Meyer, H. P. & Werbitzky, O. How Green can the industry become with biotechnology?, *in Biocatalysis for Green Chemistry and Chemical Process Development* (eds. Tao, J. & Kazlauskas, R.) 23–43 (John Wiley & Sons, Inc., Hoboken, New Jersey, USA, 2011).

29 — Meyer, H. P., Brass, J., Jungo, C., Klein, J., Wenger, J. & Mommers, R. An emerging star for therapeutic and catalytic protein production. *BioProcess International* 10–21 (2008).

30 — Ryu, Y. & Schultz, P. G. Efficient incorporation of unnatural amino acids into proteins in Escherichia coli. *Nat. Methods* **3**, 263–265 (2006).

31 — Walsh, G. Therapeutic insulins and their large-scale manufacture. *Appl. Microbiol. Biotechnol.* **67**, 151–9 (2005).

32 — Rezaei, M. & Zarkesh-Esfahani, S. H. Optimization of production of recombinant human growth hormone in Escherichia coli. *J. Res. Med. Sci.* **17**, 681–5 (2012).







33 — Sewald, N. & Jakubke, H.-D. *Peptides: Chemistry and Biology.* **3**, (Wiley-VCH Verlag GmbH & Co. KGaA, 2002).

34 — Sieber, S. A & Marahiel, M. A. Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. *Chem. Rev.* **105**, 715–38 (2005).

35 — Hodgson, D. R. W. & Sanderson, J. M. The synthesis of peptides and proteins containing non-natural amino acids. *Chem. Soc. Rev.* **33**, 422–30 (2004).

36 — van't Hoff, J. Über die zunehmende Bedeutung der anorganischen Chemie. Vortrag, gehalten auf der 70 . Versammlung der Gesellschaft deutscher Naturforscher und Ärzte zu Düsseldorf. *Zeitschrift fur Anorg. Chemie* **18**, 1–13 (1898).

37 — Ostwald, W. Die 73 . Versammlung Deutscher Naturforscher und Ärzte zu Hamburg . Über Katalyse. *Zeitschrift fur Elektrochemie* **7**, 14–16 (1901).

38 — Bergmann, M. & Fraenkel-Conrat, H. The role of specificity in the enzymatic synthesis of proteins: Syntheses with intracellular enzymes. *J. Biol. Chem.* **119**, 707–720 (1937).

39 — Zaks, A & Klibanov, A. M. Enzymatic catalysis in nonaqueous solvents. *J. Biol. Chem.* **263**, 3194–201 (1988).

40 — Klibanov, A. M. Improving enzymes by using them in organic solvents. *Nature* **409**, 241–246 (2001).

41 — Bongers, J. & Heimer, E. P. Recent applications of enzymatic peptide synthesis. *Peptides* **15**, 183–93 (1994).

42 — Bordusa, F. Proteases in organic synthesis. *Chem. Rev.* **102**, 4817–4868 (2002).

43 — Nuijens, T., Quaedflieg, P. J. L. M. & Jakubke, H. in *Enzyme Catalysis in Organic Synthesis* (eds. Drauz, K., Groger, H. & May, O.) 675–748 (Wiley-VCH Verlag GmbH & Co. KGaA, 2012).

44 — Schechter, I. & Berger, A. On the active site of proteases. Mapping the active site of papain; specific peptide inhibitors of papain. *Biochem. Biophys. Res. Commun.* **32**, 898–902 (1968).

45 — Schechter, I. Mapping of the active site of proteases in the 1960s and rational design of inhibitors / drugs in the 1990s . *Curr. Protein Pept. Sci.* **6**, 501 (2005).

46 — Didziapetris, R., Drabnig, B., Schellenberger, V., Jakubke, H. D. & Svedas, V. Penicillin acylase-catalyzed protection and deprotection of amino groups as a promising approach in enzymatic peptide synthesis. *FEBS Lett.* **287**, 31–3 (1991).

47 — Clapés, P., Torres, J. L. & Adlercreutz, P. Enzymatic peptide synthesis in low water content systems: preparative enzymatic synthesis of [Leu]- and [Met]-enkephalin derivatives. *Bioorg. Med. Chem.* **3**, 245–55 (1995).

48 — Liu, P., Tian, G., Lee, K.-S., Wong, M.-S. & Ye, Y. Full enzymatic synthesis of a precursor of bioactive pentapeptide OGP(10-14) in organic solvents. *Tetrahedron Lett.* **43**, 2423–2425 (2002).

49 — Schellenberger, V., Schwaneberg, U., Jakubke, H.-D., Hansicke, A., Bienert, M. & Krause, E. Chymotrypsin -catalyzed segment coupling synthesis of D-Phe(6)-GNRH. *Tetrahedron Lett.* **31**, 7305–7306 (1990).

50 — Nishino, N., Xu, M., Mihara, H. & Fujimoto, T. Use of hexafluoroisopropyl alcohol in tryptic condensation for partially protected precursor of α-melanocyte stimulating hormone. *Tetrahedron Lett.* **33**, 3137–3140 (1992).

51 — Koskinen, A. & Klibanov, A. M. *Enzymatic reactions in organic media.* (eds. Koskinen, A. & Klibanov, A. M. ) Blackie Academic&Professional, an imprint of Chapmann &Hall, Glasgow, UK, 1996).

52 — Nyfeler, R. in *Methods in Molecular Biology, Vol 35, Peptide Synthesis Protocols* (eds. Pennington, M. . & Dunn, B. M.) 35, 303–316 (Humana Press Inc., 1994).

53 — Dolman, M., Halling, P.J., Moore, B. D. & Waldron, S. How dry are anhydrous enzymes ? Measurement of residual and buried $^{18}$O-labeled water molecules using mass spectrometry. *Biopolymers*, **41**,313–321 (1997).

54 — Schmitke, J. L., Stern, L. J. & Klibanov, A. M. The crystal structure of subtilisin Carlsberg in anhydrous dioxane and its comparison with those in water and acetonitrile. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 4250–5 (1997).

55 — Martinek, K., Klibanov, A. M., Goldmacher, V. S. & Berezin, I. V. The principles of enzyme stabilization I. Increase in thermostability of enzymes covalently bound to a complementary surface of a polymer support in a multipoint fashion. *Biochim. Biophys. Acta - Enzymol.* **485**, 1–12 (1977).

56 — Margolin, A. L., Tai, D. & Klibanov, A. M. Incorporation of D-amino acids into peptides via enzymatic condensation in organic solvents. *J. Am. Chem. Soc.* **109**, 7885–7887 (1987).

57 — Klibanov, A. M. Why are enzymes less active in organic solvents than in water? *Trends Biotechnol.* **15**, 97–101 (1997).

58 — Nuijens, T., Cusan, C., van Dooren, T. J. G. M., Moody, H. M., Merkx, R., Kruijtzer, J. A. W., Rijkers, D. T. S., Liskamp, R. M. J. & Quaedflieg, P. J. L. M. Fully enzymatic peptide synthesis using C-terminal tert-butyl ester interconversion. *Adv. Synth. Catal.* **352**, 2399–2404 (2010).

59 — Nuijens, T., Schepers, A. H. M., Cusan, C., Kruijtzer, J. A. W., Rijkers, D. T. S., Liskamp, R. M. J. & Quaedflieg, P. J. L. M. Enzymatic fragment condensation of side chain-protected peptides using Subtilisin A in anhydrous organic solvents: A general strategy for industrial peptide synthesis. *Adv. Synth. Catal.* **355**, 287–293 (2013).

60 — Nuijens, T., Cusan, C., Kruijtzer, J. A. W., Rijkers, D. T. S., Liskamp, R. M. J. & Quaedflieg, P. J. L. M. Enzymatic synthesis of C-terminal arylamides of amino acids and peptides. *J. Org. Chem.* **74**, 5145–50 (2009).

61 — Nuijens, T., Kruijtzer, J. A. W., Cusan, C., Rijkers, D. T. S., Liskamp, R. M. J. . & Quaedflieg, P. J. L. M. A versatile and selective chemo-enzymatic synthesis of β-protected aspartic and γ-protected glutamic acid derivatives. *Tetrahedron Lett.* **50**, 2719–2721 (2009).







62 — Nuijens, T., Piva, E., Kruijtzer, J. A. W., Rijkers, D. T. S., Liskamp, R. M. J. & Quaedflieg, P. J. L. M. Fully enzymatic N→C-directed peptide synthesis using C-terminal peptide α-carboxamide to ester interconversion. *Adv. Synth. Catal.* **353**, 1039–1044 (2011).

63 — Hänsler, M. & Jakubke, H. D. Nonconventional protease catalysis in frozen aqueous solutions. *J. Pept. Sci.* **2**, 279–89 (1996).

64 — Wehofsky, N. Wespe, C., Cerovský, V., Pech, A., Hoess, E., Rudolph, R. & Bordusa, F. Ionic liquids and proteases: a clean alliance for semisynthesis. *Chembiochem* **9**, 1493–9 (2008).

65 — Nakamura, K. Biochemical reactions in supercritical fluids. *Trends Biotechnol.* **8**, 288–292 (1990).

66 — Rezaei, K., Temelli, F. & Jenab, E. Effects of pressure and temperature on enzymatic reactions in supercritical fluids. *Biotechnol. Adv.* **25**, 272–80 (2007).

67 — Krishna, S. H. Developments and trends in enzyme catalysis in nonconventional media. *Biotechnol. Adv.* **20**, 239–67 (2002).

68 — Declerck, V., Nun, P., Martinez, J. & Lamaty, F. Solvent-free synthesis of peptides. *Angew. Chemie - Int. Ed.* **48**, 9318–9321 (2009).

69 — Miyazawa, T., Ensatsu, E., Yabuuchi, N., Yanagihara, R. & Yamada, T. Superiority of the carbamoylmethyl ester as an acyl donor for the kinetically controlled amide-bond formation mediated by α-chymotrypsin. *J. Chem. Soc. Perkin Trans. 1* 390–395 (2002).

CHAPTER — 02

# Enzymes for peptide synthesis

Ana Toplak, Muhammad I. Arif, Bian Wu, and Dick B. Janssen

Biochemical Laboratory, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, 9747 AG Groningen, the Netherlands

Part of this chapter has been published in R. N. Patel, Green Biocatalysis (John Wiley & Sons, Inc., 2016)

## Abstract

Enzymes are widely used for peptide synthesis, both for coupling and for terminal activation and modification. Broad-specificity proteases of different mechanistic classes can be used for C-terminal activation and peptide ligation reactions, whereas enzymes that recognize specific sequences are increasingly used for selective modification and tagging. Protein engineering can tailor hydrolases for improved synthesis, better reaction medium tolerance and desired substrate selectivity. This chapter highlights properties of natural and engineered peptidases for chemoenzymatic peptide synthesis with emphasis on basic principles and innovations leading to new applications.



## Classification of peptidases

Peptidases suitable for peptide synthesis are found in a wide variety of organisms (Table 1). Many reports describe the use of commercially available enzymes isolated from biological sources, such as trypsin and chymotrypsin from mammalian pancreas, pepsin and chymosin from stomach, papain from plants, and subtilisin or thermolysin secreted by bacteria. Genome analysis shows that peptidases are present in all kingdoms of life: from Archaea and viruses to the higher eukaryotes. Moreover, peptidase-encoding genes comprise about 2% of the human genome, most of them having very specific functions, e.g. in hormone processing, but with unknown utility in peptide synthesis[1]. The omnipresence of peptidases, their tremendous diversity and the myriad of biological functions all contribute to the continued scientific interest in this enzyme family.

According to the enzyme classification system, peptidases belong to subgroup 4 of the hydrolases (EC 3.4.X.X). Based on the work of Rawlings and Barrett[2] the current MEROPS peptidase database counts more than 2400 peptidases and provides a description of their specificity, structural information and literature references. Peptidases are frequently classified according to their reaction mechanism, using the key groups involved in catalysis as an identifier. This gives aspartic (A), cysteine (C), glutamic (G), metallo (M), asparagine (N), serine (S), threonine (T), and unknown (U) proteases, with the letter indicated followed by an assigned number. Peptidases can also be classified in terms of families and clans, which is based on structure or sequence comparison and is a phylogenetic classification. The name of a clan is formed in most cases by a letter of the catalytic type, followed by a serial capital letter. For example, SB represents the subtilisin clan of serine peptidases and contains two families: S8 (subtilisin family) and S53 (sedolisin family). Even though members of the same clan have a common ancestor, sequence may have

diverged so much that relatedness can only be detected by structural comparison. Not surprisingly, some phylogenetic groups comprise multiple catalytic types, for which the abbreviation P was proposed. More information on peptidase classification is given by Rawlings and Salvesen[3].

One can distinguish two general types of peptidase catalytic mechanisms: catalysis via formation of a covalent intermediate and catalysis by activation of a water molecule that directly attacks the peptide bond. In case of covalent catalysis, the nucleophile attacking the peptide bond is a side chain hydroxyl or sulfhydryl, as in the serine-, threonine- and cysteine peptidases. It is a part of a classical nucleophile –base-acid catalytic triad, as in the serine hydrolases that provide the textbook example of an enzyme catalytic mechanism, or is in a functionally equivalent but structurally different variant thereof. For example, in papain the nucleophile is a cysteine occurring as a thiolate-imidazolium ion pair formed by a histidine acting as the base and a water molecule replacing the acidic group present in most serine protease catalytic triads. The formation of acyl-enzyme intermediates in serine hydrolases and related enzymes makes them suitable for application in kinetically controlled synthesis reactions (Chapter 1, Figure 5).

Metallo-, glutamic- and aspartic proteases, on the other hand, do not form covalent intermediates, but activate a water molecule that directly attacks the carbonyl carbon of the peptide bond and displaces the amide nitrogen[4]. These enzymes are often the preferred catalysts for thermodynamically controlled coupling (Chapter 1, Figure 5). An example of a metalloprotease applied for peptide synthesis is thermolysin, which contains a HExxH+E motif that coordinates a zinc ion and a water molecule. Zinc polarizes the carbonyl group and facilitates deprotonation of the water nucleophile that attacks the carbonyl group of the substrate. Catalysis is facilitated by the glutamic acid, which accepts a proton from the zinc-bound water and transfers a proton to the leaving group. In the majority of aspartic peptidases, of which gastric pepsin is the best-known example, a pair of aspartic residues act together to activate a water molecule. In glutamic peptidases, a pair of glutamic acid residues plays a similar role. The water directly attacks the peptide bond, and the reaction in peptide synthesis follows the reverse mechanism.

The serine, cysteine, aspartic and metalloproteases all have been applied in peptide synthesis. Nevertheless, shortcomings still exist and at industrial scale chemoenzymatic peptide synthesis is certainly not always the preferred method. Since the properties of the catalyst to a large extend determine the feasibility of industrial application, the discovery and engineering of better variants is an intensive field of research.





# Serine and cysteine proteases for peptide synthesis

## Chymotrypsin, trypsin and related enzymes

Classical enzymes employed for peptide coupling of the serine hydrolase family are chymotrypsin, trypsin and subtilisin. Chymotrypsin and trypsin are secreted in the mammalian gut as inactive precursors, which are activated by autoproteolysis and structural reorganization. The possibility to use chymotrypsin for peptide synthesis is known since the 1930s[5]. Most early examples concern peptide synthesis using amides or (m)ethyl esters as acyl donors, and free amino acids, their amides, short peptides, or short peptide amides as nucleophilic acyl acceptors. These studies revealed that a high pH, a high nucleophile concentration, and low product solubility stimulate formation of synthetic product. Ethyl esters appeared suitable acyl donors in kinetically controlled conversions, and amino acid amides act better as nucleophilic acyl acceptors than the free amino acids[6]. Furthermore, tripeptides often performed better than dipeptides or amino acids as acyl acceptors.

Trypsin and chymotrypsin are often used in the presence of cosolvents, which are added to increase substrate solubility, both in case of thermodynamic coupling and for kinetically controlled synthesis, and especially for longer peptides. For example, condensation of tri and heptapeptide segments, forming a $Trp^3$-$Ser^4$ linkage, in a kinetically controlled manner using α-chymotrypsin in 48% (v/v) DMF allowed over 97% conversion to a D-Phe-containing analog of gonadotropin releasing hormone ([$D\text{-}Phe^6$]GnRH) on multi-gram scale (Figure 1)[7]. Reactions with trypsins in organic solvent have also been explored. The thermodynamically controlled segment condensation of octa- and pentapeptide segments yielded an α melanocyte-stimulating hormone precursor with an $Arg^8$-$Trp^9$ ligation site. This conversion was catalyzed by trypsin in nearly dry organic solvent mixture (Figure 2)[8].

Chymotrypsin and trypsin have restricted substrate specificity. This can be an advantage since it reduces the number of possibilities for hydrolysis of a synthetic product, but also a disadvantage, since it restricts the diversity of ligation sites that can be generated. Overall, the leaving group specificity (P1', P2') of trypsin and chymotrypsin in hydrolytic reactions matches that of the nucleophile selectivity in synthetic conversions[6,9]. Chymotrypsin-catalyzed synthetic reactions are less sensitive to the nature of the acyl donor than to the nature of the nucleophile. Trypsin accepts positively charged amino acids at the P1 site (Arg, Lys) whereas chymotrypsin prefers bulky hydrophobic groups at the P1 and P1' positions.

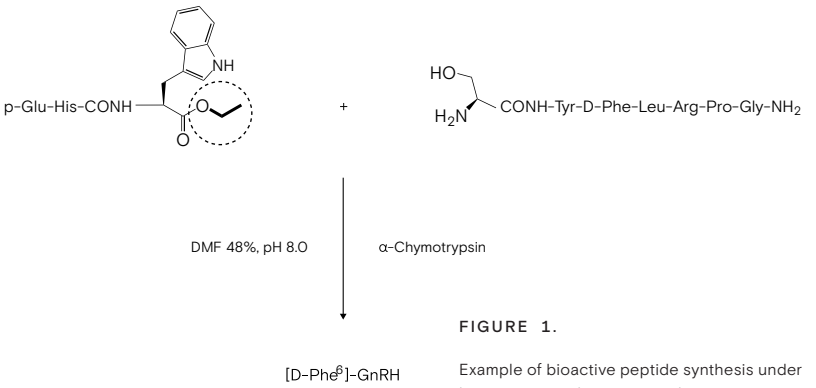



FIGURE 1.

Example of bioactive peptide synthesis under kinetic control using α-chymotrypsin in water-cosolvent medium. The gonadotropin releasing hormone analog [D-$Phe^6$]GnRH was obtained in 97.5% yield after 6.5 $h^7$.

Abbreviations: p-Glu=pyroglutamic acid.

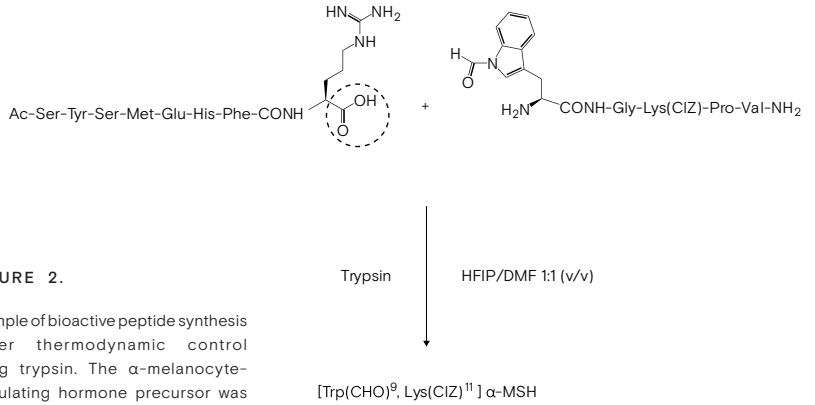



FIGURE 2.

Example of bioactive peptide synthesis under thermodynamic control using trypsin. The α-melanocyte-stimulating hormone precursor was synthetized in 95% yield using trypsin in an almost dry mixture of DMF and hexafluoroisopropylalcohol in 48 $h^8$. Protective groups: CHO = formyl group, ClZ = 2-chlorobenzyloxycarbonyl.





Proteases of the trypsin family have recently been applied for synthesis of peptides that self-assemble to nanofibrous structures and for preparing biomaterials. An example is the chymotrypsin-catalyzed polymerization of the amphiphilic peptide Lys-Leu-OEt to an alternating oligomer that forms a hydrogel[10]. Furthermore, chymotrypsin was used for the synthesis of poly-L-cysteine, a material for the selective chelation of SH-reactive metal ions under mild conditions and in high yields (80%)[11].

A trypsin-related bacterial enzyme that was tested in coupling reactions is the lysine-specific serine protease I from *Achromobacter lyticus*. The enzyme is secreted and widely used in sequence analysis of proteins[12]. It was also used in a chemoenzymatic route for the production of human insulin from porcine insulin. Since this lysyl endopeptidase cleaves only after Lys, it could be applied for replacing the C-terminus of the insulin B-chain from -Lys-Ala (porcine C-terminal sequence) to -Lys-Thr (human C-terminus) in a two-step or a single-step reaction. Using the B chain as acyl donor and Thr-O$^t$Bu in DMF-ethanol mixtures as the nucleophile, a high conversion (85-90%) was obtained[13,14]. Trypsin could also be applied in this biotransformation, but required higher enzyme loading.

## Subtilisin-like enzymes

Subtilisins have a similar catalytic mechanism as chymotrypsin but the catalytic triad is ordered in a different way along the sequence, and the enzymes are not phylogenetically related. Subtilisins are secreted by various soil-dwelling *Bacillus* strains. They comprise the largest group of commercial proteases and account for more than half of the world total sales of enzymes, mainly for use in detergents. Well-known commercial variants include subtilisin Carlsberg (also known as Alcalase), substilisin BPN' (used in many protein engineering studies), and subtilisin E. Siezen *et al.*[15] introduced the term subtilases for subtilisin-like peptidases.

Genome analysis revealed that subtilases occur in all kingdoms of life. Subtilases from the phylogenetic group of subtilisins are widely explored for the synthesis of short peptides, which is much less the case for subtilases belonging to the thermitase and proteinase K groups. However, subtilisins were reported to be less suitable for peptide synthesis than chymotrypsin and trypsin due to their high hydrolytic activity. This triggered research on subtilisin thiol variants, which can form a thioester intermediate with reduced sensitivity to hydrolysis as compared to the aminolysis reaction required for peptide coupling.

Furthermore, subtilisin has a much broader substrate range, implying that it detects a larger number of proteolytic target sites in substrates and product than chymotrypsin and trypsins.

Due to the high stability of these enzymes, peptide synthesis using subtilisin BPN' and subtilisin Carlsberg can be carried out in neat organic solvent instead of water and water-cosolvent mixtures, effectively avoiding hydrolysis of the activated precursor and of the product[16]. The use of anhydrous organic solvents also suppresses unwanted side reactions, and can influence enzyme stereoselectivity, regiospecificity, and chemoselectivity, which may allow unexpected conversion[16,17] such as coupling of activated 9-aa and 10-aa peptides carrying protecting groups from chemical peptide synthesis, resulting in the formation of a protected [19] aa peptide[18–20].

Like chymotrypsin, subtilases can catalyze reactions leading to special peptides. Proteinase K catalyzed the polymerization of L-Phe in the presence of tris(2-aminoethyl)amine to form branched oligo-(L-phenylalanine), which showed self-assembly to form fluorescent fibers[21].

## Other serine hydrolases

Of the serine exopeptidases, especially carboxypeptidase Y from *Saccharomyces cerevisiae* has been used in synthetic applications. This carboxypeptidase is an α/β-hydrolase fold enzyme, also with a Ser-Asp-His catalytic triad, but in a topology that is different from that of trypsin and subtilisin family enzymes. Carboxypeptidase Y catalyzes dipeptide coupling in aqueous medium[22]. Like with chymotrypsin, such synthetic reactions are best performed at basic pH. Protein engineering studies have revealed that the enzyme has a rather low selectivity for the P1' site, preferring hydrophobic groups, but also Lys and Arg are accepted. Selectivity can be influenced by mutations. Reducing the number of enzyme-substrate hydrogen bonds by mutating residue E65 to Ala was beneficial for the aminolysis reaction[23].





## Aminopeptidases

Prolyl aminopeptidases (PAP) are exopeptidases that hydrolytically cleave off an N-terminal Pro from peptides. The enzymes belong to the α/β hydrolase fold proteins. A prolyl aminopeptidase from *Streptomyces thermoluteus* carrying the active site mutation S144C was used as a catalyst for the synthesis of proline-containing peptides. Dipeptide synthesis was obtained with an amino acid methyl or benzyl ester as the acyl donor and prolyl-OBz as the nucleophile[24]. Under alkaline conditions cyclization and polymerization of prolyl-OBz was observed.

## Peptidases accepting ß-amino acids

In view of the growing importance of β-peptides, Kohler and coworkers explored the synthetic applicability of β-peptide specific aminopeptidases produced by strains of *Sphingosinicella* and *Ochrobactrum anthropi*. These enzymes (DmpA) belong to the P1 serine peptidase family, are distantly related to Ntn-hydrolase family, and cleave N-terminal β-amino acids from β- and α/β-peptides. In synthetic applications, oligomerization of β-amino acids occurred to produce up to octameric peptides as well as formation of α- and β-amino acid containing dipeptides with an N-terminal β-amino acid. Furthermore, the BapA enzyme from *Sphingosinicella. xenopeptidilytica* 3-2W4 could couple various β-amino acids to the N-terminus of short peptide[25]. A similar incorporation of N-terminal D-amino acids during small peptide synthesis was reported with a *Streptomyces* aminopeptidase[26].

## D-amino acid specific peptidases

Few microbial proteases acting on D-peptides are known. The alkaline D-peptidase (ADP) from *Bacillus cereus* is related to DD-carboxypeptidase and β-lactamases. These enzymes have an accessible groove in which the nucleophilic serine and other catalytic amino acids are located. This D-peptidase could be applied for the synthesis of the 92-amino-acid peptidyl prolyl *cis-trans* isomerase from *E. coli* by condensation of two peptide segments, of which the 35 amino acid acyl

donor was activated as the 4-guanidinophenyl (OGp) ester[27]. Thus the D-amino acid selective enzyme was used for preparing a protein composed of L-amino acids and making the product insensitive to hydrolysis by the coupling enzyme.

## Sulfhydryl peptidases

The use of papain in peptide synthesis is well established[28–30]. Papain can be used for the preparation of dipeptides and tripeptides, in aqueous medium with cosolvent addition (up to 40%) and at high pH to promote synthetic activity. The enzyme is a sulfhydryl protease with no homology to the trypsin or subtilase families of hydrolases. Since the catalytic nucleophile is a cysteine and because thioesters are relatively more prone to aminolysis than oxoesters, the enzyme could be very attractive for synthesis. However, unlike with the thiol variants of some serine hydrolases, the proteolytic activity is still high and the broad substrate range of proteolysis makes peptide substrate and product hydrolysis more problematic than with trypsin or chymotrypsin. Extensive enzyme engineering studies on papain are lacking, probably due to the laborious procedure for isolation of active papain from inclusion bodies formed in *E. coli*. The enzyme has a rather broad substrate range with slight preference for Lys and Arg in the S1 pocket and for hydrophobic aromatic residues in the S2 pocket. The S1' pocket is not selective which allowed the synthesis of dipeptides with non-proteinogenic amino acids[29]. To overcome the S1 preference, the substrate-mimetic approach described above was explored with the OGp leaving group in dipeptide synthesis[30]. Segment condensations with papain have also been reported, e.g. enkephalin segment coupling (2+3) was achieved in 50% yield using equimolar amounts of acyl donor (PhAc-Tyr-Gly-OMe) and nucleophile (H-Gly-Phe-Leu-O$^t$Bu) at pH 9.0 in buffer with 20% methanol[31]. Similar yields were obtained in low water systems such as buffer containing acetonitrile 4% (v/v). The synthetic potential of papain in low water systems has been explored with acyl donors bearing the carbamoylmethyl (Cam) leaving group. Dipeptide synthesis proceeded in >80% yield, and yields up to 60% were obtained for the synthesis of bioactive peptides like dermorphin-(1-4) (Boc-Tyr-D-Ala-Phe-Gly-$NH_2$). Papain is also used as a versatile protease in polymer chemistry for the synthesis of amino acid oligo- and co-oligomers of α-hydroxy acids and amino acids[32]. Homologs of the classical Papaya papain that have been tested for peptide synthesis are plant-derived enzymes such as bromelain and ficain. Bromelain extracted from pineapple could be applied in Met- and Leu-enkephalin synthesis in a low water





system, reaching 97% yield[33]. Ficain isolated from *Ficus glabrata* latex was tested in di- and tripeptide synthesis in frozen water medium[28], and was used in synthesis of Lys and Met oligomers starting from the respective ethyl esters[34].

Another papain-like enzyme that has been tested in peptide synthesis is clostripain from *Clostridium histolyticum*. It has a remarkable specificity for arginine in the S1 pocket and for proline in the S1' pocket and can be produced using a recombinant system in *E. coli* [28]. Synthesis of dipeptides such as Cbz-Arg-Pro-$NH_2$ and Cbz-Arg-D-Leu-$NH_2$ has been reported. Moreover, the substrate-mimetic approach using the OGp activating group broadened acyl donor specificity to non-proteinogenic acyl donors such as β-Ala and 4-phenylbutyric acid ester, with yields of over 86% when coupled with different nucleophiles in aqueous medium[35]. The broad specificity of the S1' pocket allowed synthesis of various peptide isosters, using Bz-Arg-OH and a range of acyl acceptor amines[36]. Animal tissue is a source of papain-related lysosomal cysteine peptidases that have been used in di- and tripeptide synthesis. Due to the insufficient homogenic preparations of cathepsin B, earlier reports of its synthetic activity have to be taken with caution[37]. The availability of a recombinant production system for human cathepsin B will allow experiments with pure enzyme. The endopeptidase cathepsin L from parasite *Fasciola hepatica* was expressed in yeast and used in the kinetically controlled synthesis of Cbz-Phe-Arg-Ser-$NH_2$ starting from Cbz-Phe-Arg-OMe and H-Ser-$NH_2$ in an aqueous medium[38].

## Sortase

For coupling of larger peptide substrates, sortases, which are responsible for covalent anchoring of surface proteins to the peptidoglycan of Gram-positive bacteria, can be employed. Sortase A (SrtA, MEROPS peptidase family C60) is a cysteine peptidase isolated from *Staphylococcus aureus* that has emerged as a powerful tool for protein bioconjugation and transpeptidation reactions[39]. During catalysis, recognition of the R-LPXT-G motif (R= protein, X=any amino acid) is followed by thiolate nucleophile attack and cleavage of the Thr-Gly amide bond. The acyl-enzyme complex (R-LPXT-SrtA) is cleaved by a peptide with an N-terminal glycine of the pentapeptide groups in peptidoglycan, yielding a covalent protein-peptidoglycan bond. This sortagging reaction can serve in applications such as protein labeling, protein-protein fusion, protein cyclization and protein immobilization. Recently, the use of sortase for the conversion

of a QALPETGEE peptide to its hydrazide derivative QALPET-$NHNH_2$ was reported. Such a peptide hydrazide can act as acyl donor in peptide coupling. Via formation of a hydrazide, this sortase reaction allowed fusion of a deoxy-D-ribose 5-phosphate aldolase segment (amino acids 1-239) with the 17 amino acid C-terminal segment, producing catalytically active enzyme connected by an LPAA linker. Similar reactions were possible with substituted hydrazines, allowing labeling of proteins containing the sortase motif with fluorescent probes[40].The efficiency of the sortase-mediated reactions depends on the flexibility and accessibility of the region comprising the recognition motif. Sortase engineering and mutant selection for better motif recognition has been reported[41]. Details regarding sortase mechanism and applications of the enzyme can be found in an excellent review by M. Ritzefeld[39].

## Metalloproteases in peptide synthesis

### Thermolysin-like enzymes

The metalloprotease thermolysin has been widely used for peptide synthesis[42,43]. Thermolysin is produced and secreted by the thermophilic bacterium *Bacillus thermoproteolyticus*. Catalysis involves water activation by a zinc ion that is coordinated by side chains of two histidines and a glutamate. Thermolysin specificity is defined predominantly by its S1' pocket that accepts large hydrophobic residues, but also polar and charged residues. In the S1 pocket hydrophobic residues are preferred, and Leu>Ala>Phe>Gly is the preference order for the S2 and S2' pockets. This hydrolytic specificity of thermolysin is reflected in its synthetic specificity. Many related so-called thermolysin-like proteinases (TLPs) from various Gram-positive strains have been described[3], including neutral proteases from *Bacillus subtilis*, and some of these variants are applied in peptide synthesis. Several metalloenzymes acting as carboxy- or aminopeptidase have also been characterized but these variants have not been extensively used in peptide synthesis. A bovine carboxypeptidase A[44] and orange carboxypeptidase C[45] have been applied for dipeptide synthesis in water-organic solvent mixtures, both under thermodynamic and under kinetic control.





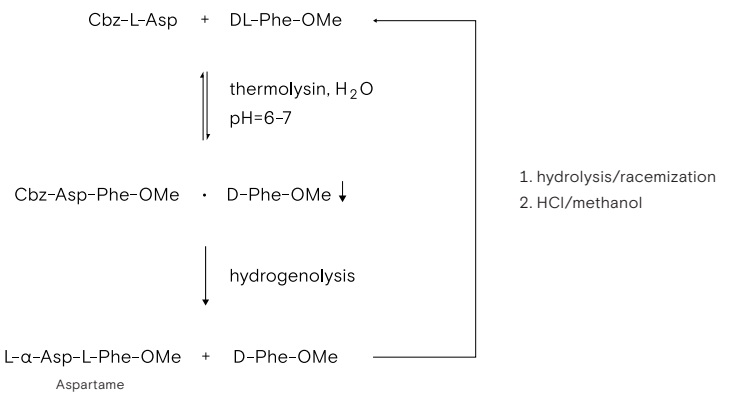



FIGURE 3.

Chemoenzymatic peptide synthesis of aspartame using thermolysin under thermodynamic control. The coupling reaction is stereo- and regioselective. The unreacted isomer D-Phe-OMe forms a precipitate with the product, shifting the equilibrium toward the synthesis. After precursor isolation and hydrogenolysis, the D-isomer is chemically racemized and can be re-used[46,47].

With thermolysin and related metalloproteases, peptide synthesis is carried out under thermodynamic control where the reversal of hydrolysis is achieved by modifying reaction conditions in such a way that the equilibriums shifts to synthesis. Thermolysin has been employed in numerous di- and tripeptide coupling reactions i.e. for the synthesis of artificial sweetener aspartame (L-α-Asp-L-Phe-OMe) on multi-ton level. Unlike the chemical route, which requires optically pure starting materials and gives substantial formation of byproducts, the enzymatic synthesis is an elegant and environmental friendly method. The reaction is performed under mild conditions and is both regioselective (only the α-carboxyl group reacts) and stereoselective (for L-Phe), thus allowing the use of racemic acyl acceptor (D,L-Phe-OMe). The D-isomer of Phe forms a precipitate with Cbz-L-α-Asp-L-Phe-OMe, thereby changing the equilibrium in favor of synthesis. After deprotection, D-Phe can be racemized and recycled (Figure 3). This is an example of substrate-induced precipitation for thermodynamic control of peptide synthesis.

Another example of thermolysin-catalyzed peptide synthesis is the production of precursors of enkephalins[16]. Also coupling of non-natural amino acids, such as halophenylalanines, is possible[48]. Similar to α-chymotrypsin and papain, thermolysin has been used for peptide polymerization[49]. Using the reversibility of thermolysin-catalyzed peptide bond formation, dynamic combinatorial libraries of peptides could be established and screened for the formation of nanostructures with special properties[50].

## Aspartic proteases in peptide synthesis

Aspartic proteases form a group of proteolytic enzymes that catalyze peptide bond cleavage by acid-base catalysis and activation of a water molecule for nucleophilic attack on the amide carbon. Crystal structures of mammalian and fungal enzymes are known. In pepsin, the best studied aspartic protease, catalysis proceeds by water activation and leaving group protonation. Both involve an aspartate, which explains the low pH optimum of around 4. The two aspartic acid residues are situated around a hydrophobic cleft that can accommodate 7 amino acids (the S4-S3' subsites). The active site is covered by a flexible flap, which contributes to S1 subsite specificity.

Since the catalytic cycle does not include an acyl-enzyme intermediate, attempts to use pepsin for synthesis have been focused on thermodynamically controlled conversions[51]. Conversions can be done with addition of organic solvent to increase medium hydrophobicity and shift the equilibrium to synthesis. Synthesis of several tri- and hexapeptides by pepsin in buffer containing 18% DMF reached yields of 30-78%. Pepsin can also be used in two-phase systems. The synthetic yields are determined by enzyme specificity, equilibrium of the coupling reactions, as well as by partitioning of substrates and products over the two phases. Pepsin has a substrate preference for hydrophobic peptides but in coupling reactions of Cbz-Xaa-Phe-OH and Phe-OMe a better yield was observed with hydrophilic groups in the P2 position[52,53]. Hydrophobic peptides and some non-proteinogenic substrates are well accepted in the S1' pocket, which may render pepsin a useful catalyst for the synthesis of non-natural peptides[51]. In case immobilized enzyme is used, it is preferable to keep the substrate and products in solution to avoid problems with separation of products from the catalyst. For this reason, DMF has been added as a solubilizing agent, benefiting from the fact that pepsin is highly resistant to denaturing conditions.





A prominent aspartic protease related to pepsin is chymosin, produced in the calf stomach for casein hydrolysis. Its use as a catalyst in peptide synthesis has been described, again in thermo-dynamically catalyzed coupling reactions, where linkages are introduced in hydrophilic regions of the target peptides. Carboxybenzoyl (Cbz) and *p*-nitroanilides were used as amine- and acyl-protecting groups, respectively. Using DMF to keep the substrates in solution and acetate buffer at low pH (5.3) to stimulate synthesis, coupling products were formed and precipitated[53]. This way, several tetra- to hexapeptides were produced. The recombinant overexpression of chymosin in *E. coli*, yeast and, for industrial production, in filamentous fungi is well established. Consequently, chymosin variants could be obtained by protein engineering, and mutants with higher pH optimum and higher thermostability have been described[54]. The use of these enzymes in peptide synthesis remains underexplored.

**TABLE 1.**

Peptidases used for chemoenzymatic enzymatic peptide synthesis.

| ENZYME | SOURCE | SPECIFICITY | APPLICATION |
|---|---|---|---|
| **Serine peptidases** | | | |
| α-chymotrypsin<br>clan PA<br>EC 3.4.21.1 | pancreas<br>recombinant yeast system available for rat enzyme | P1=Tyr, Trp, Phe, Leu<br>P1' nonspecific | used in synthesis of bioactive peptides in water-organic solvent mixtures and ionic liquids; substrate mimetic mediated coupling of long segments[6–9,55]; amino acid polymers for biomaterials [10] |
| trypsin<br>clan PA<br>EC 3.4.21.4 | bovine, porcine, murine pancreas<br>recombinant systems: *E. coli, P. pastoris, S. cerevisiae* | P1=Arg, Lys<br>P1' nonspecific | synthesis of bioactive peptides in water and water-cosolvent; insulin transpeptidation8; substrate mimetic approach[56] |
| subtilisins<br>clan SB<br>EC 3.4.21.62 | various *Bacillus sp.,*<br>recombinant production in *E. coli* | P1=large, non-branched hydrophobic residues preferred<br>P1'= Gly > Ser, Lys, Ala > Arg > Gln, Thr, Val, Met, Asn > His > Ile, Trp > Glu > Asp, Leu.Pro | bioactive peptide synthesis in neat organic solvent and water-organic solvent mixtures;<br>amino acid oligomerization[18-20]; engineered thiol variants for enhanced synthesis[57,58] |
| proteinase K<br>clan SB<br>EC 3.4.21.64 | secreted from fungus *Tritrachium album* | P1=aromatic, hydrophobic residues<br>P1'=similar to subtilisins | [Leu] enkephalin peptide synthesis in organic solvent in a multistage reactor[59]; synthesis of branched peptides[21] |
| protease V8<br>clan PA<br>EC 3.4.21.19 | extracellular peptidase from *Staphylococcus aureus* | P1=Asp, Glu | substrate mimetic mediated coupling of dipeptides and peptide segments[60,61] |
| carboxypeptidase Y<br>clan SC<br>EC 3.4.16.5 | production in yeast | P1'=hydrophobic residues | C-terminal extension in water-organic solvent medium[23] |
| prolyl aminopeptidase (PAP)<br>clan SC<br>EC 3.4.11.5 | *Streptomyces* sp. | P1 = Pro, P1'= broad | Synthesis of peptides and short polymers containing Pro |
| **Cysteine peptidases** | | | |
| apain,<br>clan CA<br>EC 3.4.22.2 | papaya plant latex(*Carica papaya*), recombinant production in *S. cerevisiae* and *E. coli* | P1=Arg, Lys; P2=aromatic hydrophobic<br>P1'= large hydrophobic | kinetically controlled peptide synthesis; synthesis of bioactive peptides in organic solvent (demorphin, enkephalin); amino acid oligomerization[32]; di- and tri-peptide synthesis with peptidomimetic leaving group[28–31,33] |

| ENZYME | SOURCE | SPECIFICITY | APPLICATION |
|---|---|---|---|
| bromelain,<br>clan CA<br>EC 3.4.22.33 | pineapple fruit *(Ananas comosus)* | P1 and P1'=polar amino acids | [Met]- and [Leu]-enkephalin segment condensation (2+3)[33] |
| ficain,<br>clan CA<br>EC 3.4.22.3 | extraction from *Ficus glabrata* latex | P1= Gly, Ser, Glu, Tyr, Phe; P2=hydrophobic side chains | Lys-Met ethyl ester oligomer synthesis; short peptide synthesis in frozen aqueous medium[34] |
| clostripain,<br>clan CD<br>EC 3.4.22.8 | *Clostridium histolyticum,*<br>recombinant production in *E. coli* and *Bacillus subtilis* | P1=Arg, P1'= D-amino acids, Pro | di-and tripeptide synthesis in frozen aqueous medium; synthesis of peptide isosters[35] |
| cathepsin L,<br>clan CA<br>EC 3.4.22.15 | *Fasciola hepatica,* recombinant production in yeast | P2=large hydrophobic residues, P1'= Ser > Ala > Lys > Asn > Gln | Cbz-Phe-Arg-Ser-$NH_2$ synthesis under kinetic control[38] |
| sortase A,<br>clan CL<br>EC 3.4.22.70 | *Staphylococcus aureus, E. coli* expression system available | LPXT-G, X= any amino acid | peptide and protein labeling, cyclization and immobilization, protein-protein fusion, cell surface labeling[39,62] |
| **Metallo peptidases** | | | |
| thermolysin<br>clan MA<br>zinc endopeptidase<br>EC 3.4.24.27 | secreted by *Bacillus thermoproteolyticus*,<br>recombinant in *E. coli* | P1' = large hydro¬pho¬bic, also polar and charged residues,<br>P1= hydrophobic residues, P2, P2'= Leu > Ala > Phe > Gly, P3'=basic residues | aspartame precursor synthesis, dipeptide synthesis[48,63,64] |
| stearolysin<br>clan MA<br>zinc endopeptidase | secreted by *Geobacillus stearothermophilus,*<br>recombinant production in *E.coli* | similar to thermolysin, P1'=Phe | thermostable variant boilysin in dipeptide Cbz-Asp-X-OMe (X= Phe, D-Phe, Ala, Ile, Leu, Met, Tyr or Val) synthesis[64] |
| pseudolysin<br>(elastase),<br>clan MA<br>EC 3.4.24.26 | ecreted by *Pseudomonas aeruginosa*,<br>recombinant production in *E.coli* and *Pichia* | P1'=hydrophobic or aromatic residues;Phe > Leu > Tyr > Val, Ile > Ala<br>P1,P2'=Ala | dipeptide synthesis, including substrates as norvaline and norleucine, Cbz-Ala-Phe-$NH_2$ synthesis[65] |
| vimelysin,<br>clan MA | secreted by *Vibrio* sp. T1800; recombinant production in *E.coli* | P1'=Phe | aspartame precursor synthesis in the presence of 30% DMSO and ethanol[66] |
| carboxypeptidase A<br>clan MC<br>EC 3.4.17.1 | *Bos taurus;* recombinant production in yeast | P1'=large, hydrophobic residues | dipeptide synthesis in low water systems, 27% highest yield obtained[44] |
| D,D-dipeptidase | *Brevibacillus borstelensis* BCS-1;<br>recombinant production in *E.coli* | P1'=D-amino acid | Cbz-L-Asp-D-AlaOBzl alitame sweetener precursor synthesis[67] |
| aminopeptidase SSAP | *Streptomyces septatus,* recombinantly produced in *E. coli* | broad specificity, inactive towards N-protected peptides | various dipeptide methyl esters starting from free acyl donor and aminoacid methyl ester as nucleophile in methanol [24] |
| **Aspartic peptidases** | | | |
| pepsin<br>clan AA<br>EC 3.4.23.1 | porcine, recombinant production in *E. coli* | P1 and P1'=hydrophobic residues,<br>P2=hydrophilic residues | various di- and -tri-peptides in two phase systems, hexapeptide synthesis at pH 4.6 in 20% (*v/v*) DMF[51-53] |
| chymosin<br>clan AA<br>EC 3.4.23.4 | calf stomach (bovine, camel), recombinant production in: *E. coli,* yeast and filamentous fungi | P1 and P1'=hydrophobic residues | tetra- and hexapeptide synthesis at pH 5.3 in 20% *v/v* DMF[54] |

## Protease discovery

The restricted applicability of commercially available proteases for peptide coupling reactions warrants the continuous importance of mining natural resources for enzymes with novel properties, such as a broad substrate range, acceptance of non-proteinogenic amino acids, high coupling efficiency, and robustness under reaction conditions.

### Metagenomics

With current protocols less than 1% of the existing microorganisms can be isolated in the laboratory, therefore metagenomic approaches comprising cloning and sequence- or function-based screening can contribute to rapid biocatalyst discovery. Genes can be discovered that are very different from those with established function. For example, using metagenomics a metallo-protease was discovered that has less than 30% similarity to any known protease[68]. The potential of metagenomics was recently also demonstrated for variants of subtilisin. In total 51 different subtilisins were discovered, each carrying 2-8 substitutions that are distant from the active site but affect temperature stability and substrate specificity[69]. It seems likely that these approaches can discover variants with altered characteristics in peptide synthesis.

### Proteases from thermophiles

Whereas subtilisin Carlsberg, subtilisin BNP' and subtilisin E display moderate temperature stabilities, several thermostable proteases of the subtilase superfamily have been obtained from extremophilic organisms. Examples are aqualysin I from *Thermus aquaticus* and thermitase from *Thermoactinomyces vulgaris*. Although over 30 of such characterized thermophilic peptidases are known, only few have been applied in peptide synthesis. The serine peptidase form *Thermus aquaticus* Rt41A (PreTaq) was applied in kinetically controlled synthesis of Bz-Ala-Tyr-$NH_2$ with modest yield (26%) using immobilized enzyme in DMF (90% v/v), Bz-Ala-OMe as acyl donor and a large excess of Tyr-$NH_2$ as





the nucleophile at 40°C and pH 10[70]. Two serine peptidases of the subtilase family from the hyperthermophilic organisms *Thermus aquaticus* and *Deinococcus* geothermalis were also applied in peptide synthesis under harsh conditions such as high temperature (60 and 80°C) in anhydrous solvent with addition of DMF as cosolvent (40% v/v) (Chapter 5).

### Solvent-tolerant proteases

The hypothesis that extracellular enzymes of organic-solvent tolerant microorganisms should exert resistance to such organic solvents led to the discovery of a solvent-tolerant metallo-protease called pseudolysin produced by the bacterium *Pseudomonas aeruginosa*. Organic solvent tolerance of this metallo-protease was attributed to a disulfide bond (Cys30-Cys58) and to the presence of charged residues at the enzyme surface (Y45, N201). In the thermodynamic approach using medium with 50% DMF, the enzyme could effectively synthetize the dipeptide Cbz-Arg-Leu-NH2 (78% yield), and the aspartame precursor Cbz-Asp-Phe-OMe (89% yield) in the presence of 50% DMSO[71]. Similarly, the opioid peptides endomorphin-1 (Tyr-Pro-Trp-Phe-$NH_2$)[72] and endomorphin-2 (Tyr-Pro-Phe-Phe-$NH_2$)[73] were synthesized by the organic-solvent-tolerant metallo-proteases from *Pseudomonas aeruginosa* and *Bacillus cereus*, respectively. Both peptidases tolerated up to 50% (v/v) of DMSO and DMF and conversions yielded endomorphin-2 with 80% overall yield. The se enzymes are all structurally related to thermolysin. Recently, a novel subtilisin-like serine peptidase has been discovered in *Pseudomonas mendocina* ymp by genome mining and described (Chapter 6).

### Proteases from salt-resistant organisms

Halophilic microorganisms that evolved to survive low water activity conditions are promising for discovery of peptidases that survive under harsh conditions. A halophilic serine peptidase from *Halobacterium salinarum* was reported to catalyze synthesis of di- and tripeptides, e.g. producing Ac-Phe-Gly-Leu-$NH_2$ form Ac-Phe-OEt as acyl donor in 76% yield in the presence of 33% (v/v) DMF, 20 equivalents of the nucleophile dipeptide amide and 2.8 M NaCl[74].

The acyl donor specificity of the enzyme is quite broad, but the nucleophile preferred at S1' is Gly, Ala or another small hydrophobic group. Although promising, the synthetic potential of halophilic peptidases is still poorly exploited, and biochemical information on such enzymes is scarce.

## Proteases engineered for improved synthesis

Protein engineering of enzymes for peptide synthesis has mainly focused on improving the synthesis:hydrolysis ratio of serine hydrolases used in kinetically controlled peptide synthesis and on studies aimed at re-engineering enzyme specificity. Other targets are improving the resistance to organic solvents and tailoring the hydrolytic activity, but only a few of these engineered enzymes were tested later as catalysts in peptide coupling.

### Solvent-resistant and thermostable subtilase mutants

Extensive work on stabilizing subtilisins by mutagenesis has been carried out. Thermostability and organic cosolvent tolerance are often positively correlated. Stabilizing mutations include introduction of groups that form salt bridges, expanding the internal hydrogen bonding, improving hydrophobic interactions, and internal crosslinking by disulfide bonds[75–77]. An excellent overview on subtilisin BPN' engineering is given by P. Bryan[78]. The synthetic potential of a DMF-tolerant subtilisin BPN' mutant was tested. This enzyme, subtilisin 8350, carries mutations N218S (improved hydrogen bonding), G169A (improved hydrophobic interaction and conformational restriction), M50F (improved hydrophobic interaction), Q206C (better van der Waals interactions), N76D (improved $Ca^{2+}$ binding and hydrogen bonding) and Y217K (better H bonding). The enzyme catalyzed synthesis of several bioactive peptides in 50% DMF under kinetic control, including enkephalin release factor (Tyr-D-Arg, 70% yield), chemotactic peptide (Met-Leu-Phe-$NHCH_2C_6H_5$, 95% yield), [Leu]-enkephalinamide and dermorphin in yields above 70%. The enzyme also catalyzed oligomerization of Met-OMe (formation of a 50-mer) where the degree of polymerization was influenced by the amount of DMF added[79].





An even more stable variant of subtilisin BPN', subtilisin 8397 (M50F,N76D,G169A, Q206C,N218S) was designed and used in glycopeptide synthesis[57] and the synthesis of a peptide heptamer in 50% DMF[80]. Further mutagenesis to increase the stability in the presence of DMF revealed K43N and K256Y as important mutations. In addition, the mutations M222A and Y217W broadened the P1' specificity of this subtilisin. The authors also report regioselective transesterification, transacylation, and amidation of peptide esters[81].

### Thermostable thermolysin variants

Thermolysin has been subjected to extensive protein engineering studies aimed at enhancing thermostability[82]. Such studies yielded mutants (D150A, D150E, D150W, I168A, and N227H) that had improved activity in aspartame synthesis as compared to the wild type[63]. Engineering the thermostability of stearolysin, a protease from *Geobacillus stearothermophilus* CU21 that is 86% sequence identical to thermolysin, resulted in an extremely stable 8-fold variant that was active at 100°C and in the presence of denaturants. This enzyme, termed boilysin, had comparable specificity as thermolysin in synthesis of dipeptides (Cbz-Asp-X-OMe (X= Phe, D-Phe, Ala, Ile, Leu, Met, Tyr or Val)) and increased activity at elevated temperatures and at low calcium concentrations[64].

### Increasing aminolysis to hydrolysis ratios by protein engineering

Protein engineering methods aimed at increasing the aminolysis to hydrolysis ratio of peptidases have focused on replacing the active site serine by a cysteine. Early studies on thiol-subtilisin[83] and seleno-subtilisin[58] demonstrated increased aminolysis to hydrolysis ratio in peptide synthesis in aqueous medium or in DMF-containing buffers. Thiol- subtilisin is a weak peptidase, and can only rapidly cleave esters. Since thio- and selenol-acyl-enzyme intermediates undergo aminolysis faster than oxo-esters, these enzymes are particularly useful for the synthesis of amide bonds. However, the esterolytic activity is lower than that of the wild-type enzyme and the modified enzymes are oxidant-sensitive. Mutagenesis of subtilisin BPN' with concomitant introduction of the mutations S221C and P225A yielded a variant more suitable for peptide synthesis[84].

This variant of subtilisin mutant was termed subtiligase. It forms acyl-enzyme intermediates only with acyl donors activated as esters, and the glycolate phenylalanyl amide functionality was selected as an activating group. The specificity of subtiligase has been studied, but the ligation efficiency is still not absolutely predictable, as it does not directly reflect subsite specificities[85]. By screening tripeptides as acyl acceptor, it was found that especially the P1' specificity influences ligation efficiency. The potential of subtiligase is clear from the chemoenzymatic synthesis of intact and catalytically active ribonuclease A (124 amino acids) by sequential ligation of six segments in the C→N direction with 10% overall yield on milligram scale[86]. Further mutagenesis of subtiligase included the introduction of stabilizing mutations (M50F, N76D, N109S, K213R, N218), yielding a denaturant-resistant variant[87]. In addition, random mutagenesis at 25 positions around the active site of subtiligase and screening the coupling activity of a phage-displayed enzyme library yielded variants with modified selectivity. By screening for attachment of a labeled peptide to its N-terminus, two subtiligase mutants were identified (M124L,L126V and M124L, S125A) with improved ligase activity[88]. The applicability of subtiligase is broad, and it has been used for selective labeling of proteins by coupling a peptide to the N-terminus of a target protein[89]. Subtiligase has also been used for kinetically controlled peptide backbone cyclization. The cyclization of peptides of length 12-31 amino acids activated as the C-terminal OCam phenylalanylamide ester was performed in aqueous solution[90].

A peptide ligase, termed peptiligase has been recently reported (Chapter 7). This highly engineered, thermostable and $Ca^{2+}$ independent subtilisin BPN' variant, carries the S221C and P225A mutations that resulted in improved synthetic properties over state of the art enzyme subtiligase.

## Protein engineering of trypsin-like proteases

Variants of trypsin with increased aminolysis rate in substrate-mimetic mediated peptide synthesis have been obtained by mutating active site residues and residues in the S1 subsite. A variant called trypsiligase was described by Bordusa and coworkers[91]. It carries the mutations D189K (reduces Arg and tolerates Tyr specificity in the S1 subsite), K60E (introduces Arg selectivity in the S1' subsite) and N143H, E151H (introduces Zn-dependent His specificity in the S2' subsite), and thus recognizes the Y↓RH sequence motif.





The enzyme could be applied to process proteins carrying an YRH- linker at the N-terminus, and the resulting RH- N-terminus could be labeled by acting as a nucleophilic substrate in tryptiligase acyl-enzyme formed from an OGp-activated acyl donor[84]. Trypsiligase adopts a partially disordered inactive zymogen-like conformation, and is activated by addition of the peptide substrate and zinc ions. Trypsiligase has also been applied in C-terminal protein modification, e.g. PEG- or dye-labeling of a protein C-terminus containing trypsiligase recognition YRH motif[92]. A Fab segment derived from a therapeutic antibody used in breast cancer treatment was extended at the C-terminus of its heavy chain with a tag containing the YRH- recognition sequence. This was followed by functionalization using PEG- or dye-labeled RHAK peptide as a nucleophile in a transpeptidation reaction. Trypsiligase-mediated protein modification offers an alternative to other enzyme-mediated peptide modification reactions, such as those catalyzed by subtiligase, sortase A, and transglutaminase.

Mutagenesis studies on the chymotrypsin-like protease B from *Streptomyces griseus* (SGBP) revealed ligation activity when the active site serine was replaced by Ala or Gly (S195A/G)[93]. The ligation is proposed to proceed through a histidine-involved acyl-enzyme intermediate. The stability of the enzyme was improved by introducing mutations with stabilizing effects (S195G and T213L), yielding streptoligase[94] .

## Computational design

Alongside structure-based engineering and directed evolution, computational protein design is becoming a powerful tool for tailoring enzymes to specific biotechnological applications. Recently, a *de novo* designed peptide ligase based on a 33 α-helical peptide that forms a coiled coil, has been reported[95]. Furthermore, a peptidase with designed specificity towards α-gliadin has been developed using computational methods by D. Baker and coworkers[96]. The rapid developments in this field suggest that computational methods will soon contribute to the redesign of peptidases for synthetic applications.

## Conclusions

Success in chemoenzymatic peptide synthesis and peptide modification is critically dependent on the selection of correct enzyme-substrate combinations. Many classical proteases have been successfully used in peptide synthesis, and key factors are medium engineering and substrate engineering. The latter includes selection of coupling sites for enzyme-mediated segment condensation, and selection of leaving groups in kinetically controlled synthesis. Strongly activating leaving groups that allow formation of high-energy intermediates that yield stable products in subsequent aminolysis reactions expand the application scope. Many reactions can be done in the presence of cosolvents to increase peptide solubility or in neat organic solvent to suppress peptide hydrolysis. This requires highly stable enzymes, which have been obtained by protein engineering. Neat organic solvents may change enzyme selectivity, and broaden the range of reactions that can be performed. When working in aqueous conditions, the use of thiol-subtilisins is advantageous and allows efficient segment coupling.

**Author contribution**

AT, MIA, BW and DBJ wrote the manuscript.





## References


01 — Marnett, A. B. & Craik, C. S. Papa's got a brand new tag: advances in identification of proteases and their substrates. *Trends Biotechnol.* **23**, 57–9 (2005).

02 — Rawlings, N. D., Barrett, A. J. & Bateman, A. MEROPS: the peptidase database. *Nucleic Acids Res.* **38**, D227–33 (2010).

03 — Rawlings, N. D. & Salvesen, G. *Handbook of proteolytic enzymes.* (Academic Press, Elsevier, 2013).

04 — Lipscomb, W. N. & Straeter, N. Recent advances in zinc enzymology. *Chem. Rev.* **96**, 2375–2433 (1996).

05 — Bergmann, M. & Fraenkel-Conrat, H. The role of specificity in the enzymatic synthesis of proteins: Syntheses with intracellular enzymes. *J. Biol. Chem.* **119**, 707–720 (1937).

06 — Schellenberger, V., Schellenberger, U., Mitin, Y. V. & Jakubke, H.-D. Characterization of the S'-subsite specificity of bovine pancreaticα-chymotrypsin via acyl transfer to added nucleophiles. *Eur. J. Biochem.* **187**, 163–167 (1990).

07 — Schellenberger, V., Schwaneberg, U., Jakubke, H.-D., Hansicke, A., Bienert, M. & Krause, E. Chymotrypsin-catalyzed fragment coupling synthesis of D-Phe(6)-GNRH. *Tetrahedron Lett.* **31**, 7305–7306 (1990).

08 — Nishino, N., Xu, M., Mihara, H. & Fujimoto, T. Use of hexafluoroisopropyl alcohol in tryptic condensation for partially protected precursor of α-melanocyte stimulating hormone. *Tetrahedron Lett.* **33**, 3137–3140 (1992).

09 — Schellenberger, V., Braune, K., Hoffmann, H.-J. & Jakubke, H. The specificity of chymotrypsin. *Eur. J. Biochem.* **636**, 623–636 (1991).

10 — Qin, X., Xie, W., Tian, S., Cai, J., Yuan, H., Yu, Z., Butterfoss, G. L., Khuong, A. C. & Gross, R. A. Enzyme-triggered hydrogelation via self-assembly of alternating peptides. *Chem. Commun. (Camb).* **49**, 4839–41 (2013).

11 — Narai-Kanayama, A., Hanaishi, T. & Aso, K. α-Chymotrypsin-catalyzed synthesis of poly-L-cysteine in a frozen aqueous solution. *J. Biotechnol.* **157**, 428–36 (2012).

12 — Kishimoto, T., Kondo, J., Takai-Igarashi, T. & Tanaka, H. Accurate mass comparison coupled with two endopeptidases enables identification of protein termini. *Proteomics* **11**, 485–9 (2011).

13 — Morihara, K. Enzymatic semisynthesis of human insulin: An update. *J. Mol. Recognit.* **3**, 181–186 (1990).

14 — Morihara, K., Oka, T., Tsuzuki, H., Tochino, Y. & Kanaya, T. *Achromobacter* protease I-catalyzed conversion of porcine insulin into human insulin. *Biochem. Biophys. Res. Commun.* **92**, 396–402 (1980).

15 — Siezen, R. J. & Leunissen, J. A. Subtilases: the superfamily of subtilisin-like serine proteases. *Protein Sci.* **6**, 501–23 (1997).

16 — Koskinen, A. & Klibanov, A. M. *Enzymatic reactions in organic media.* (eds. Koskinen, A. & Klibanov, A. M. ) Blackie Academic&Professional, an imprint of Chapmann &Hall, Glasgow, UK, 1996).

17 — Margolin, A. L., Tai, D. & Klibanov, A. M. Incorporation of D-amino acids into peptides via enzymatic condensation in organic solvents. *J. Am. Chem. Soc.* **109**, 7885–7887 (1987).

18 — Nuijens, T., Cusan, C., van Dooren, T. J. G. M., Moody, H. M., Merkx, R., Kruijtzer, J. A. W., Rijkers, D. T. S., Liskamp, R. M. J. & Quaedflieg, P. J. L. M. Fully enzymatic peptide synthesis using C-terminal tert-butyl ester interconversion. *Adv. Synth. Catal.* **352**, 2399–2404 (2010).

19 — Nuijens, T., Schepers, A. H. M., Cusan, C., Kruijtzer, J. A. W., Rijkers, D. T. S., Liskamp, R. M. J. & Quaedflieg, P. J. L. M. Enzymatic fragment condensation of side chain-protected peptides using Subtilisin A in anhydrous organic solvents: A general strategy for industrial peptide synthesis. *Adv. Synth. Catal.* **355**, 287–293 (2013).

20 — Nuijens, T., Piva, E., Kruijtzer, J. A. W., Rijkers, D. T. S., Liskamp, R. M. J. & Quaedflieg, P. J. L. M. Fully enzymatic N→C-directed peptide synthesis using C-terminal peptide α-carboxamide to ester interconversion. *Adv. Synth. Catal.* **353**, 1039–1044 (2011).

21 — Ageitos, J. M., Baker, P. J., Sugahara, M. & Numata, K. Proteinase K- catalyzed synthesis of linear and star oligo(L-Phenylalanine) conjugates. *Biomacromolecules* **14**, 3635–3642 (2013).

22 — Christensen, U. Kinetic characterization of carboxypeptidase Y-catalyzed peptide semisynthesis. Prediction of yields. *Amino Acids* **6**, 177–187 (1994).

23 — Mortensen, U. H., Stennicke, H. R., Raaschou-Nielsen, M. & Breddam, K. Mechanistic study on carboxypeptidase Y-catalyzed transacylation reactions. Mutationally altered enzymes for peptide synthesis. *J. Am. Chem. Soc.* **116**, 34–41 (1994).

24 — Yamamoto, Y., Usuki, H., Iwabuchi, M. & Hatanaka, T. Prolyl aminopeptidase from *Streptomyces thermoluteus* subsp. *fuscus* strain NBRC14270 and synthesis of proline-containing peptides by its S144C variant. *Appl. Environ. Microbiol.* **76**, 6180–5 (2010).

25 — Heck, T., Kohler, H.-P. E., Limbach, M., Flogel, O., Seebach, D. & Geueke, B. Enzyme-catalyzed formation of β-peptides : β-peptidyl aminopeptidases BapA and DmpA acting as β-peptide-synthesizing enzymes aminopeptidases from proteobacteria are able to cleave β-peptides.*Helv. Chim. Acta* **4**, 2016–2030 (2007).




CHAPTER — 02 — REFERENCES


26 — Arima, J., Morimoto, M., Usuki, H., Mori, N. & Hatanaka, T. ß-Alanyl peptide synthesis by *Streptomyces* S9 aminopeptidase. *J. Biotechnol.* **147**, 52–58 (2010).

27 — Wehofsky, N., Pech, A., Liebscher, S., Schmidt, S., Komeda, H., Asano, Y. & Bordusa, F. D-amino acid specific proteases and native all-L-proteins: a convenient combination for semisynthesis. *Angew. Chem. Int. Ed. Engl.* **47**, 5456–60 (2008).

28 — Hänsler, M., Ullmann, G. & Jakubke, H. D. The application of papain, ficin and clostripain in kinetically controlled peptide synthesis in frozen aqueous solutions. *J. Pept. Sci.* **1**, 283–7

29 — Barbas, C. & Wong, C. Papain catalysed peptide synthesis: Control of amidase activity and the introduction of unusual amino acids. *J. Chem. Soc. Chem. Commun*. 533–534 (1987).

30 — de Beer, R. J. A. C., Zarzycka, B., Mariman, M., Amatdjais-Groenen, H., Mulders, M. J., Quaedflieg, P. J. L. M., van Delft, F. L., Nabuurs, S. B. & Rutjes, F. P. J. T. Papain-specific activating esters in aqueous dipeptide synthesis. *Chembiochem* **13**, 1319–26 (2012).

31 — Didziapetris, R., Drabnig, B., Schellenberger, V., Jakubke, H. D. & Svedas, V. Penicillin acylase-catalyzed protection and deprotection of amino groups as a promising approach in enzymatic peptide synthesis. *FEBS Lett.* **287**, 31–3 (1991).

32 — Schwab, L. W., Kloosterman, W. M. J., Konieczny, J. & Loos, K. Papain catalyzed (co)oligomerization of α-amino acids. *Polymers (Basel).* **4**, 710–740 (2012).

33 — Clapés, P., Torres, J. L. & Adlercreutz, P. Enzymatic peptide synthesis in low water content systems: preparative enzymatic synthesis of [Leu]- and [Met]-enkephalin derivatives. *Bioorg. Med. Chem.* **3**, 245–55 (1995).

34 — Andre, M., Kühl, B., Brenner-Weiss, G., Syldatk, C. & Rudat, J. Cationic heterooligopeptides by ficain-catalyzed co-oligomerization of lysine and methionine ethylesters. *J. Pept. Sci.* **20**, 625–9 (2014).

35 — Guenther, R., Stein, A. & Bordusa, F. Investigations on the enzyme epecificity of Clostripain : A new efficient biocatalyst for the synthesis of peptide isosteres. *J. Org. Chem.* **65**, 1672–1679 (2000).

36 — Günther, R. & Bordusa, F. Protease catalysis mediated by a substrate mimetic: a novel enzymatic approach to the synthesis of carboxylic acid amides. *Chemistry* **6**, 463–7 (2000).

37 — Fruton, J. S. Proteinase-catalyzed synthesis of peptide bonds. *Adv. Enzymol. Relat. Areas Mol. Biol.* **53**, 239–306 (1982).

38 — Ruth, D. M., McMahon, G. & O'Fágáin, C. Peptide synthesis by recombinant *Fasciola hepatica* cathepsin L1. *Biochimie* **88**, 117–20 (2006).

39 — Ritzefeld, M. Sortagging: a robust and efficient chemoenzymatic ligation strategy. *Chemistry* **20**, 8516–29 (2014).

40 — Li, Y.-M. Li, Y.-T., Pan, M., Kong, X.-Q., Huang, Y.-C., Hong, Z.-Y. & Liu, L. Irreversible site-specific hydrazinolysis of proteins by use of sortase. *Angew. Chemie* **126**, 2230–2234 (2014).

41 — Chen, I., Dorr, B. M. & Liu, D. R. A general strategy for the evolution of bond-forming enzymes using yeast display. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 11399–404 (2011).

42 — Oka, T. & Morihara, K. Peptide bond catalyzed by thermolysin. *J. Biochem.* **88**, 807–813 (1980).

43 — Wayne, S. I. & Fruton, J. S. Thermolysin-catalyzed peptide bond synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **80**, 3241–3244 (1983).

44 — Vértesi, A. & Simon, L. M. Carboxypeptidase A-catalyzed dipeptide synthesis in organic media. *J. Biotechnol.* **66**, 75–82 (1998).

45 — Steinke, D., Schwarz, a, Wandrey, C. & Kula, M. R. Formation of peptide bonds by carboxypeptidase c from orange leaves. *Enzyme Microb. Technol.* **13**, 262–6 (1991).

46 — Quaedflieg, P. J. L. M., Sonke, T. & Wagner, A. F. V. Synthesis and recovery of aspartame involving enzymatic deformylation step.US patent 6,617,127 B2 filed 22 Jun. 2001, and issued 9 Sept. 2003.

47 — Bommarius, A. S. & Riebel-Bommarius, B. *Biocatalysis: Fundamentals and Applications.* (Wiley-VCH Verlag GmbH & Co. KGaA, 2004).

48 — Imaoka, Y., Kawamoto, T., Ueda, M. & Tanaka, A. Peptide synthesis with halophenylalanines by thermolysin. *Appl. Microbiol. Biotechnol.* **40**, 653–656 (1994).

49 — Williams, R. J., Smith, A. M., Collins, R., Hodson, N., Das, A.K., & Ulijn, R. V. Enzyme-assisted self-assembly under thermodynamic control. *Nat. Nanotechnol.* **4**, 19–24 (2009).

50 — Nalluri, S. K. M. & Ulijn, R. V. Discovery of energy transfer nanostructures using gelation-driven dynamic combinatorial libraries. *Chem. Sci.* **4**, 3699 (2013).

51 — Bemquerer, M. P., Adlercreutz, P. & Tominaga, M. Pepsin-catalyzed peptide synthesis in organic media: studies with free and immobilized enzyme. *Int. J. Pept. Protein Res.* **44**, 448–456 (1994).

52 — Malak, C. Filippova, I. Y., Lysogorskaya, E. N., Animisova, V. V., Lavrenova, G. I. & Stepanov, V. M. Pepsin as a catalyst of peptide synthesis. Enzyme co-precipitation with emerging peptide products. *Int. J. Pept. Protein Res.* **39**, 443–449 (1992).

53 — Malak, C. Pepsin as a catalyst for peptide synthesis : formation of peptide bonds not typical for pepsin substrate specificity. *J. Pept. Res.* **53**, 606–610 (1999).

54 — Abdel Malak, C. A. Calf chymosin catalyst of peptide synthesis. *J. Biochem.* **288**, 941–943 (1992).







55 — Morihara, K. & Oka, T.α-chymotrypsin as the catalyst for peptide synthesis. *J. Biochem.* **163**, 531–542 (1977).

56 — Bordusa, F., Ullmann, D., Elsner, C. & Jakubke, H. Substrate mimetic mediated peptide synthesis:An irreversible ligation strategy that is independent of substrate specificity. *Angew. Chem. Int. Ed. Engl.* **4**, 2473–2475 (1997).

57 — Wong, C., Schuster, M., Wang, P. & Searst, P. Enzymatic synthesis of N- and O-linked glycopeptides. *J. Am. Chem. Soc.* **115**, 5893–5901 (1993).

58 — Wu, Z. & Hilvert, D. Conversion of a protease into an acyl transferase: Selenolsubtilisin. 4513–4514 (1989).

59 — Fruton, J. S. Proteinases as catalysts of peptide bond synthesis. *Trans. New York Acad. Sci.* 49–56 (1972).

60 — Wehofsky, N., Wissmann, J., Alisch, M. & Bordusa, F. Engineering of substrate mimetics as novel-type substrates for glutamic acid-specific endopeptidases : design, synthesis , and application. *Biochem. Biophys. Acta* **1479**, 114–122 (2000).

61 — Wehofsky, N. & Bordusa, F. Programming of enzyme specificity by substrate mimetics : investigations on the Glu-specific V8 protease reveals a novel general principle of biocatalysis. *FEBS Lett.* **443**, 220–224 (1999).

62 — Wu, Z., Guo, X. & Guo, Z. Sortase A-catalyzed peptide cyclization for the synthesis of macrocyclic peptides and glycopeptides. *Chem. Commun. (Camb).* **47**, 16–19 (2011).

63 — Kusano, M., Yasukawa, K. & Inouye, K. Synthesis of N-carbobenzoxy-l-aspartyl-l-phenylalanine methyl ester catalyzed by thermolysin variants with improved activity. *Enzyme Microb. Technol.* **46**, 320–325 (2010).

64 — Kühn, D., Dürrschmidt, P., Mansfeld, J. & Ulbrich-Hofmann, R. Boilysin and thermolysin in dipeptide synthesis: a comparative study. *Biotechnol. Appl. Biochem.* **36**, 71–76 (2002).

65 — Rival, S., Saulnier, J. & Wallach, J. On the mechanism of action of pseudolysin: Kinetic Study of the enzymatic condensation of Z-Ala with Phe-$NH_2$. *Biocatal. Biotransformation* **17**, 417–429 (2000).

66 — Kunugi, S., Koyasu, A., Takahashi, S. & Oda, K. Peptide condensation activity of a neutral protease from *Vibrio* sp. T1800 (Vimelysin ). *Biotechnol. Bioeng.* **53**, 387–390 (1997).

67 — Baek, D. H., Song, J. J., Kwon, S.-J., Jung, C.-M., Sung, M.-H. & Park, C. Characteristics of a new enantioselective thermostable dipeptidase from *Brevibacillus borstelensis* BCS-1 and its application to synthesis of a D-amino-acid-containing dipeptide. *Appl. Environ. Microbiol.* **70**, 1570–1575 (2004).

68 — Gupta, R., Beg, Q. K. & Lorenz, P. Bacterial alkaline proteases: molecular approaches and industrial applications. *Appl. Microbiol. Biotechnol.* **59**, 15–32 (2002).

69 — Gabor, E., Niehaus, F., Aehle, W. & Eck, J. Zooming in on metagenomics: molecular microdiversity of subtilisin Carlsberg in soil. *J. Mol. Biol.* **418**, 16–20 (2012).

70 — Wilson, S. A., Daniel, R. M. & Peek, K. Peptide synthesis with a proteinase from the extremely thermophilic organism *Thermus* Rt41A. *Biotechnol. Bioeng.* **44**, 337–46 (1994).

71 — Tsuchiyama, S., Doukyu, N., Yasuda, M., Ishimi, K. & Ogino, H. Peptide synthesis of aspartame precursor using organic-solvent-stable PST-01 protease in monophasic aqueous-organic solvent systems. *Biotechnol. Prog.* **23**, 820–823 (2007).

72 — Sun, H., He, B., Xu, J., Wu, B. & Ouyang, P. Efficient chemo-enzymatic synthesis of endomorphin-1 using organic solvent stable proteases to green the synthesis of the peptide. *Green Chem.* **13**, 1680 (2011).

73 — Xu, J., Sun, H., He, X., Bai, Z. & He, B. Highly efficient synthesis of endomorphin-2 under thermodynamic control catalyzed by organic solvent stable proteases with in situ product removal. *Bioresour. Technol.* **129**, 663–6 (2013).

74 — Ryu, K., Kim, J. & Dordick, J. S. Catalytic properties and potential of an extracellular protease from an extreme halophile. *Enzyme Microb. Technol.* **16**, 266–75 (1994).

75 — Almog, O. Gallagher, D. T., Ladner, J. E., Strausberg, S., Alexander, P., Bryan, P. & Gilliland, G. L. Structural basis of thermostability. Analysis of stabilizing mutations in subtilisin BPN'. *J. Biol. Chem.* **277**, 27553–8 (2002).

76 — Zhao, H. & Arnold, F. H. Directed evolution converts subtilisin E into a functional equivalent of thermitase. *Protein Eng. Des. Sel.* **12**, 47–53 (1999).

77 — Chen, K. Q. Robinson, A. C., Van Dam, M. E., Martinez, P., Economou, C. & Arnold, F. H. Enzyme engineering for nonaqueous solvents. II. Additive effects of mutations on the stability and activity of subtilisin E in polar organic media. *Biotechnol. Prog.* **7**, 125–129 (1991).

78 — Bryan, P. N. Protein engineering of subtilisin. *Biochim. Biophys. Acta* **1543**, 203–222 (2000).

79 — Wong, C.-H., Chen, S., Hennen, W. J., Bibbs, J. A., Wang, Y., Liu, J. L., Pantoliano, M. W., Whitlow, M. & Bryan, P. N. Enzymes in Organic Synthesis: Use of subtilisin and a highly stable mutant derived from multiple site-specific mutations. *J. Am. Chem. Soc.* **112**, 945–953 (1990).

80 — Zhong, Z. Liu, J. L., Dinterman, L. M., Finkelman, M. A. J., Mueller, W. T., Rollence, M. L., Whitlow, M. & Wong, C.-H. Engineering subtilisin for reaction in dimethylformamide. *J. Am. Chem. Soc.* **100**, 683–684 (1991).

81 — Sears, P., Schuster, M., Wang, P., Witte, K. & Wong, C.-H. Engineering subtilisin for peptide coupling: Studies on the effects of counterions and site-specific modifications on the stability and specificity of the Enzyme. *J. Am. Chem. Soc.* **116**, 6521–6530 (1994).

82 — van den Burg, B., Vriend, G., Veltman, O., Venema, G. & Eijsink, V. Engineering an enzyme to resist boiling. *Proc. Natl. Acad. Sci.* **95**, 2056–2060 (1998).

83 — Nakatsuka, T., Sasaki, T. & Kaiser, E. Peptide segment coupling catalyzed by the semisynthetic enzyme thiolsubtilisin. *J. Am. Chem. Soc.* **11**, 3808–3810 (1987).







84 — Abrahmsén, L., Tom, J., Burnier, J., Butcher, K. A., Kssiakoff, A. & Wells, J. A. Engineering subtilisin and its substrates for efficient ligation of peptide bonds in aqueous solution. *Biochemistry* **30**, 4151–9 (1991).

85 — Braisted, A., Judice, J. & Wells, J. A. Synthesis of proteins by subtiligase. *Methods Enzymol.* **289**, 298–313 (1997).

86 — Jackson, D. Y. Burnier, J., Quan, C., Stanley, M., Tom J. & Wells, J. A. A designed peptide ligase for total synthesis of ribonuclease A with unnatural catalytic residues. *Science* **266**, 243–247 (1994).

87 — Chang, T. K., Jackson, D. Y., Burniert, J. P. & Wells, J. A. Subtiligase : A tool for semisynthesis of proteins. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 12544–12548 (1994).

88 — Atwell, S. & Wellst, J. A. Selection for improved subtiligases by phage display. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 9497–9502 (1999).

89 — Yoshihara, H. A. I., Mahrus, S. & Wells, J. A. Tags for labeling protein N-termini with subtiligase for proteomics. *Bioorg. Med. Chem. Lett.* **18**, 6000–3 (2008).

90 — Jackson, D. Y., Bumier, J. P. & Wellst, J. A. Enzymatic cyclization of linear peptide esters using subtiligase. *J. Am. Chem. Soc.* **117**, 819–820 (1995).

91 — Liebscher, S., Schöpfel, M., Aumüller, T., Sharkhuukhen, A., Pech, A., Höss, E., Parthier, C., Jahreis, G., Stubbs, M. T. & Bordusa, F. N-terminal protein modification by substrate-activated reverse proteolysis. *Angew. Chem. Int. Ed. Engl.* **53**, 3024–8 (2014).

92 — Liebscher, S., Kornberger, P., Fink, G., Trost-Gross, E.-M., Höss, E., Skerra, A & Bordusa, F. Derivatization of antibody fab fragments: a designer enzyme for native protein modification. *Chembiochem* **15**, 1096–100 (2014).

93 — Elliott, R. J., Bennet, A. J., Braun, C. A., MacLeod, A. M. & Borgford, T. J. Active-site variants of *Streptomyces griseus* protease B with peptide-ligation activity. *Chem. Biol.* **7**, 163–171 (2000).

94 — Joe, K., Borgford, T. J. & Bennet, A. J. Generation of a thermostable and denaturant-resistant peptide ligase. *Biochemistry* **43**, 7672–7 (2004).

95 — Kennan, A. J., Haridas, V., Severin, K., Lee, D. H. & Ghadiri, M. R. A de novo designed peptide ligase : A mechanistic investigation. *J. Am. Chem. Soc.* **123**, 8282–8286 (2001).

96 — Gordon, S. R. Stanley, E. J., Wolf, S., Toland, A., Wu, S. J., Hadidi, D., Mills, J. H., Baker, D., Swanson Pultz, I. & Siegel, J. B. Computational design of an α-gliadin peptidase. *J. Am. Chem. Soc.* **134**, 20513–20 (2012).

CHAPTER — 03

# Genome mining for novel proteases

Ana Toplak and Dick B. Janssen

Biochemical Laboratory, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, 9747 AG Groningen, the Netherlands

Part of this chapter has been published in: Toplak A, Wu B, Fusetti F, Quaedflieg PJML, Janssen DB. Proteolysin, a novel highly thermostable and cosolvent-compatible protease from the thermophilic bacterium *Coprothermobacter proteolyticus*. Appl Environm Microbiol (2013) 79: 5625-5632.

## Abstract

Enzymes from extremophiles offer the opportunity to expand the range of reaction conditions for biocatalytic conversions. The wealth of microbial genome sequence information allows the exploration for genes encoding proteins with expected high thermostability and/or organic solvent tolerance. In this work, we report the cloning and screening for functional expression in *E. coli* of robust subtilases identified by genome mining.

## Introduction

Extreme ecosystems in terms of temperature, pressure, salinity and pH represent a valuable resource of microorganisms that produce exquisite enzymes[1]. The high stability of such thermozymes (enzymes from extreme thermophiles) is of great importance for developing new industrial applications[2,3]. Thermozymes not only show resistance to high temperatures, but often also tolerate harsh conditions such as the presence of denaturants and organic solvents[4]. Industrial processes that operate under these conditions benefit from enhanced substrate solubility, lowered liquid viscosity, increased diffusion rates of substrates and products, as well as from favorable equilibrium shifts in certain endothermal reactions. In addition, thermostable enzymes tend to be more resistant to proteolysis[5], and may be compatible with heat-treatment aimed at reducing microbial contamination[6,7]. Robust microorganisms in terms of resistance to harsh environments are also important for the biodegradation of environmental pollutants. Such bioremediation organisms are used for sanitation of contaminated areas (heavy metals, organic solvents, radioactive compounds) and might also be a source of valuable enzymes.

Subtilisins are serine proteases from *Bacillus* strains; they comprise the largest group of commercial proteolytic enzymes[8] and account for more than half of the world total sales of enzymes[9]. They are extensively used in food, textile, detergent, pharmaceutical and leather industries[10–13]. A few thermostable subtilisin homologs (subtilases) have been isolated and characterized, for example from Archaea (e.g. pyrolysin[14], stetterlysin[15], pernisine[16], Tk-SP[17], Tk-1689 and Tk-subtilisin[18,19]), and from thermophilic bacteria (fervidolysin[20], Ak.1 protease[21], Rt41A proteinase[22,23], aqualysin I[24], islandisin[25] and thermitase[26]). In order to discover novel thermostable subtilases that resist harsh process conditions, we explored available genome sequences from thermophiles, hyperthermophiles and some bioremediation organisms for homologs of subtilisin E.

Since cultivation of extremophiles is associated with potential difficulties, cloning and expression of protease genes into a mesophilic host that is easy to grow (*E. coli, B. subtilis*, yeast) is of importance for biochemical investigation, protein engineering studies and practical enzyme production. Another possible advantage of using a mesophilic host is that enzyme isolation can be based on the difference in thermostability between host proteins and the target thermozyme[27]. Unfortunately, expression levels in mesophilic hosts using standard expression systems rarely exceed a few mg per L of culture. In case of proteases, autocatalytic processing and secretion may be specific bottlenecks[19,28]. Most subtilases are synthesized as a precursor with a pre- and a prosequence which are present as an N-terminal extension of the mature catalytic protein. Occasionally, there is a C-terminal extension as well. The N-terminal presequence serves as a signal peptide for the translocation through the cell membrane. By serving as intramolecular chaperone and inhibitor of the proteolytic activity the prosequence is involved in folding of the mature enzyme[28,29]. The C-terminal prodomain is involved in secretion to the extracellular medium in the original host[29,30], whereas in *E. coli* it facilitates the translocation of the enzyme to the outer membrane. It can be removed during a heating step, as shown for aqualysin I[31,32].

Here, we describe the genome mining for novel extremophilic subtilases, their cloning, and functional expression of 10 novel subtilases in *E. coli*.





## Materials and methods

**Chemicals and reagents —** Restriction enzymes were from New England BioLabs (Beverly, MA, USA). The PCR system was supplied by Finnzymes (Vantaa, Finland), and the LigaFast ligase system and trypsin were from Promega (Madison, WI, USA). All chemicals and subtilisin A type VIII (Carlsberg, Alcalase) were purchased from Sigma-Aldrich (St. Louis, MO, USA).

**Bacterial strains and plasmids —** Genomic DNA of *Thermus aquaticus* Y51MC23 was from Lucigen (Middleton, WI, USA). *Pseudomonas mendocina* strain ymp genomic DNA was a kind gift from Prof. J.L. DuBois (Notre Dame University, IN, USA) and genomic DNA of *Geobacillus thermodenitrificans* NG80-2 was a kind gift from Prof. X. Liu (Nankai University, Tianjin, China). Cells of *Deinococcus geothermalis* (DSM11300) and *C. proteolyticus* (DSM5265) were purchased from DSMZ (Braunschweig, Germany). *E. coli* strain C43(DE3) was used for cloning and expression and was from Lucigen (Middleton, WI, USA). Cloning plasmids included pBADMycHisA (Invitrogen, San Diego, CA, USA), pBADMycHisA(NdeI) with the NcoI site of pBAD mutated to an NdeI site, pET28a+, pET20b+ (Novagen, Merck KGaA, Darmstadt, Germany) and pBAD-MBP, a derivative of plasmid pBADMycHisA(NdeI) with an additional maltose binding protein (MBP) gene to which the 5'end of the target gene is connected via a linker; and optional C-terminal hexahistidine tag also connected via a linker.

**Cloning of the protease genes —** The genes for putative extremophilic proteases were cloned using either cells or genomic DNA as the template. CloneManager 7 (Sci-Ed Software) was used to design PCR primer pairs. In primer design the stop codon was omitted or replaced by another amino acid, and the start codon was modified to ATG, if necessary. A PCR protocol for GC-rich sequences was established: the 50 µl reaction mixture contained 2 µl template DNA, 1x GC-buffer (Finnzymes), 0.2 mM deoxynucleoside triphosphates (New England Biolabs), 0.4 µM of each primer with 0.02 U/µl of Phusion polymerase (Finnzymes) and 3% DMSO. Amplification was performed with a thermocycler (PEQlab) programmed for 5 min at 98°C initial denaturation, followed by 35 cycles, each cycle consisting of 30 s at 98°C, 30 s at 50-68°C, 45 s at 72°C, with a final 10 min extension at 72°C. The PCR products were ligated in the respective restriction enzyme-treated vectors (pBADMycHisA, pBADMBP, pET28a+ or pET20b+) and transformed to *E. coli* C43(DE3) (see Table 1 and Table 2).

The sequence of the cloned genes was confirmed by DNA sequencing (GATC, Konstanz, Germany).

**Cloning of the proteolysin gene —** The gene coding for the complete preprosequence of proteolysin (EMBL accession no. YP_002247839) was amplified by whole-cell PCR on *C. proteolyticus* cells using the forward primer 5'-GCCCGCGCCATATGAAAAAGATACTATTAACACTGGTTATCG-3' (NdeI site underlined) and the reverse primer 5'-CACACACGAAGCTT**TTA**TGGTGTCCAGTTTACTGCAGCATAC-3' (HindIII site underlined, stop codon changed to Leu codon in bold). The PCR products were ligated in NdeI- and HindIII-treated pBAD-MBP vector, and transformed to *E. coli* C43(DE3). This yielded a recombinant plasmid termed pBAD-MBP-PrlA encoding the maltose binding protein, the signal peptide, the N-terminal prosequence, the mature protein and a putative C-terminal region ending with a linker (LKLGPEQKLISEEDLNSAVD) and a hexahistidine tag. The genes for other putative extremophilic proteases were cloned in a similar manner using either cells or genomic DNA as the template (Table 3.2 and 3.3). The sequence of the cloned genes was confirmed by DNA sequencing (GATC, Konstanz, Germany).

**Activity screening on plates —** For testing protease expression, transformants were grown on LB-agar plates (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, 15g/L agar) containing 1% skim milk, appropriate antibiotic and inducer (0.025% L-(+) arabinose or 0.2 mM IPTG (isopropyl β-D-1-thiogalactopyranoside)) for 48 h at 30°C. Activity was apparent from halo formation around colonies. Colonies were also subjected to heat treatment (1-2 h at 80°C) to identify bacteria expressing proteases that require high-temperature maturation.

**Expression, processing and isolation of subtilases —** *E. coli* C43(DE3) cells containing an expression construct were grown overnight at 37°C in LB medium supplemented with 0.1% glucose and ampicillin (100 µg/ml). The overnight culture was used to inoculate (0.1% v/v) 50-100 ml TB medium (12 g/L tryptone, 24 g/L yeast extract, 0.4% (v/v) glycerol, 72 mM $K_2HPO_4$, 17 mM $KH_2PO_4$) containing ampicillin (100 µg/ml) or kanamycin (50 µg/ml) and cells were grown at 4 different temperatures (17, 24, 30 and 37°C) for various times (24 h for 37 and 30°C and 48 h for 24 and 17°C). Subtilase synthesis was induced when the OD600 reached 0.8 by adding 0.25% arabinose or 0.3 mM IPTG, followed by cultivation with shaking. The cells were harvested (4°C, 15 min at 6,000 x g),





washed, suspended in buffer (50 mM Hepes·NaOH, pH 7.5, 10 mM $CaCl_2$), and disrupted by sonication. The total cell lysate was divided and subjected to heat treatment at two temperatures (50°C and 70°C). Samples were taken after 2 h and 20 h of incubation.

**Qualitative protease activity determination —** Proteolytic activities of cell extracts and purified enzyme samples were monitored on 5% skim milk LB agar plates. Samples were added to wells in the agar and proteolytic activity was scored after overnight incubation at 37°C by inspection for clearing zonesHepes·NaOH, pH 7.5, 10 mM $CaCl_2$), and disrupted by sonication. The total cell lysate was divided and subjected to heat treatment at two temperatures (50°C and 70°C). Samples were taken after 2 h and 20 h of incubation.

# Results and discussion

## Genome mining

In the search for the homologs of subtilisin E in extremophiles, the whole preprosequence of subtilisin E (P04189) was used as a query in BLASTp searches. The target database comprised of sequenced hyperthermophiles (Archaea), thermophiles (Bacteria) and bioremediation organisms from *Pseudomonas* and *Deinococcus* genera (Bacteria) available in December 2008. Through genome mining we identified 73 hits (BLASTp S-value ≥ 80-cut off) from 36 different organisms (Table 1). In detail, 13 hits were found in 9 archaeal organisms and 60 hits in 27 bacterial strains. From this number, unavailable organisms and known proteases reported by other groups were dismissed resulting in 50 remaining potential hits (6 hits from 6 different Archaea and 44 bacterial hits from 18 organisms).The selected genes were PCR-amplified from cells or from isolated genomic DNA. In total 23 genes coding for putative proteases from 18 different microorganisms were inserted into in several cloning vectors using restriction cloning. Constructs were transformed into the cloning strain *E. coli* TOP10 and into *E.coli* C43(DE3) as the expression strain (Table 2 and Table 3).





| BACTERIA | NUMBER OF HITS |
|---|---|
| *Acidothermus cellulolyticus* 11B | 1 |
| *Anaerocellum thermophilum* DSM 6725 | 1 |
| *Anoxybacillus flavithermus* WK1 | 2 |
| *Clostridium thermocellum* DSM 4150 | 1 |
| *Coprothermobacter proteolyticus* DSM 5265 | 1 |
| *Deinococcus geothermalis* DSM 11300 | 4 |
| *Deinococcus radiodurans* R1 | 4 |
| *Fervidobacterium nodosum* Rt17-B1 | 3 |
| *Geobacillus kaustophilus* HTA426 | 2 |
| *Geobacillus sp.* G11MC16 | 3 |
| *Geobacillus sp.* WCH70 | 3 |
| *Geobacillus sp.* Y412MC10 | 3 |
| *Geobacillus thermodenitrificans* NG80-2 | 3 |
| *Halothermothrix orenii* H 168 | 4 |
| *Heliobacterium modesticaldum* Ice1 | 1 |
| *Natranaerobius thermophilus* JW/NM-WN-LF | 2 |
| *Peudomonas mendocina* ymp | 1 |
| *Symbiobacterium thermophilum* IAM 14863 | 3 |
| *Thermoanaerobacter pseudethanolicus* ATCC 33223 | 1 |
| *Thermoanaerobacter sp.* X514 | 2 |
| *Thermoanaerobacter tengcongensis* MB4 | 2 |
| *Thermobifida fusca* YX | 1 |
| *Thermosipho africanus* TCF52B | 2 |
| *Thermosipho melanesiensis* BI429 | 1 |
| *Thermotoga lettingae* TMO | 1 |
| *Thermotogales bacterium* TBF 19.5.1 | 1 |
| *Thermus aquaticus* Y51MC23 | 3 |
| *Thermus thermophilus* HB8 | 4 |
| 27 strains | 60 |

| ARCHAEA | NUMBER OF HITS |
|---|---|
| *Archaeoglobus fulgidus* DSM 4304 | 1 |
| *Desulfurococcus kamchatkensis* 1221n | 1 |
| *Methanosaeta thermophila* PT | 1 |
| *Pyrobaculum aresenaticum* DSM13514 | 1 |
| *Pyrobaculum islandicum* DSM4184 | 1 |
| *Pyrococcus furiosus* DSM 3638 | 1 |
| *Thermococcus kodakarensis* KOD1 | 3 |
| *Thermococcus onnurineus* NA1 | 3 |
| *Thermofilum pendens* Hrk5 | 1 |
| 9 strains | 13 |

TABLE 2.

Cloning of putative proteases with ≥ 21% sequence identity to subtilisin E (P04189) originating from thermophilic and radiophilic Bacteria.

* activity observed after 1-2 h additional heating step at 80°C ; NT, not tested

| Accession number | Organism | Source | % Sequence Identity to subtilisin E | Restriction Sites | Vector | Expression host | Plate activity | Flask expression | Enzyme name, Ref. |
|---|---|---|---|---|---|---|---|---|---|
| YP_001127191.1 | *Geobacillus thermodenitrificans* NG80-2 | gDNA | 39 | NcoI-ApaI | pBADMycHis (NdeI) | *E. coli* C43(DE3) | active | inactive | |
| YP_001189588.1 | *Pseudomonas mendocina* ymp | gDNA | 36 | NdeI-HindIII | pBADMycHis (NdeI)<br>pET28a+ | *E. coli* C43(DE3) | active | confirmed,<br>best results obtained<br>using pBADMycHis (NdeI) | *Pm*Sbt<br>Chapter 6 |
| YP_002247839.1 | *Coprothermobacter proteolyticus* | DSM5265 | 35 | NdeI-HindIII | pBADMycHis (NdeI)<br>pBADMBP<br>pET28a+ | *E. coli* C43(DE3) | active | confirmed,<br>best results obtained<br>using pBADMBP | proteolysin<br>Chapter 4 |
| NP_294536.1 | *Deinococcus radiodurans* R1 | DSM20539 | 35 | NdeI-HindIII | pBADMycHis (NdeI) | *E. coli* C43(DE3) | inactive | inactive | |
| YP_001125483.1 | *Geobacillus thermodenitrificans* NG80-2 | gDNA | 32 | NdeI-HindIII | pBADMycHis (NdeI)<br>pBADMBP<br>pET28a+ | *E. coli* BL21 (DE3) | inactive | active | NG483Sbt<br>this chapter |
| ZP_01188835.1 | *Halothermothrix orenii* H168 | DSM9562 | 32 | NdeI-HindIII | pBADMycHis (NdeI) | *E. coli* C43(DE3) | inactive | NT | |
| YP_604447.1 | *Deinococcus geothermalis* | DSM11300 | 31 | NdeI-HindIII | pBADMycHis (NdeI) | *E. coli* C43(DE3) | active | confirmed | *Dg*Sbt Chapter 5 |
| YP_001471655.1 | *Thermotoga lettingae* TMO | DSM14385 | 30 | NdeI-HindIII | pET20b+ | *E. coli* C43(DE3) | inactive | NT | |
| ZP_03495941.1 | *Thermus aquaticus* Y51MC23 | Lucigen, gDNA | 30 | NdeI-HindIII | pBADMycHis (NdeI)<br>pET28a+ | *E. coli* C43(DE3) | active | confirmed,<br>best results obtained<br>using pET28a+ | *Taq*Sbt<br>Chapter 5 |
| ZP_03327874.1 | *Anaerocellum thermophilum* | DSM 6725 | 26 | NcoI-ApaI | pBADMycHis (NdeI) | *E. coli* C43(DE3) | inactive | NT | |
| NP_285606.1 | *Deinococcus radiodurans* R1 | DSM 20539 | 26 | NdeI-HindIII | pBADMycHis (NdeI) | *E. coli* C43(DE3) | inactive | inactive | |
| YP_001124317.1 | *Geobacillus thermodenitrificans* NG80-2 | gDNA | 25 | NdeI-HindIII | pBADMBP<br>pET28a+ | *E. coli* C43(DE3) | active | confirmed,<br>best results obtained<br>using pET28a+ | NG317Sbt<br>this chapter |
| ZP_01189676.1 | *Halothermothrix orenii* H168 | DSM 9562 | 28 | NdeI-HindIII | pBADMycHis (NdeI) | *E. coli* C43(DE3) | inactive | NT | |
| YP_001665698.1 | *Thermoanaerobacter pseudethanolicus* | DSM 2355 | 28 | NdeI-HindIII | pBADMycHis (NdeI) | *E. coli* C43(DE3) | active* | NT | |
| YP_288282.1 | *Thermobifida fusca* YX | gDNA | 21 | NdeI-HindIII | pBADMycHis (NdeI) | *E. coli* C43(DE3) | inactive | NT | |
| ZP_03496757.1 | *Thermus aquaticus* Y51MC23 | gDNA | 23 | NdeI-HindIII | pBADMycHis (NdeI) | *E. coli* C43(DE3) | active* | confirmed | |
| ZP_03496880.1 | *Thermus aquaticus* Y51MC23 | gDNA | 22 | NdeI-HindIII | pBADMycHis (NdeI) | *E. coli* C43(DE3) | active* | confirmed | |
| YP_143552.1 | *Thermus thermophilus* HB8 | DSM 579 | 24 | NdeI-HindIII | pBADMBP | *E. coli* C43(DE3) | inactive | NT | |

TABLE 3.

Cloning of putative proteases with ≥ 29% sequence identity to subtilisin E (P04189) originating from hyperthermophilic Archaea.

NT, not tested

| Accession number | Organism | Source | % Sequence Identity to subtilisin E | Restriction Sites | Vector | Expression host | Plate activity | Flask expression |
|---|---|---|---|---|---|---|---|---|
| YP_001154254.1 | *Pyrobaculum aresenaticum* | DSM 13514 | 37 | NdeI-EcoRI | pBADMBP | *E. coli* C43(DE3) | inactive | NT |
| NP_579399.1 | *Pyrococcus furiosus* | DSM 3638 | 37 | NdeI-EcoRI | pBADMBP | *E. coli* C43(DE3) | inactive | inactive |
| YP_931297.1 | *Pyrobaculum islandicum* | DSM 4184 | 36 | NdeI-HindIII | pBADMBP | *E. coli* C43(DE3) | inactive | inactive |
| NP_070481.1 | *Archaeoglobus fulgidus* | DSM 4304 | 39 | NdeI-HindIII | pBADMBP | *E. coli* C43(DE3) | inactive | NT |
| YP_843693.1 | *Methanosaeta thermophila* PT | DSM 6194 | 29 | NdeI-HindIII | pBADMycHis(NdeI) | *E. coli* C43(DE3) | inactive | NT |

Transformed *E. coli* strains expressing active subtilases were identified on 1% skim milk plates containing an appropriate inducer and antibiotic. Some of the putative proteases were cloned in different vectors to identify a generic plasmid for subtilase expression. A set of 6 strains with different inserts showed activity after 48 h incubation at 30°C on skim milk plates (Table 2 and 3, column 8). Since processing of thermophilic proteases may require an elevated temperature as shown for aqualysin I[32] plates were incubated at 80°C for additional 1-2 h after the initial scoring for protease production. Three additional strains were identified. One more subtilase was identified in a small-scale liquid culture test. Thus, in total 10 genes encoding active protease were identified.

Strains with proteolytic activity on plate were tested for expression and activity in liquid cultures (Table 2 and 3, column 9). Typical expression tests involved cultivation at 4 different temperatures (17, 24, 30 and 37°C) with 0.25% arabinose or 0.3 mM IPTG as inducer. To allow autocatalytic processing of the precursor of thermophilic proteases, protease activities were assayed on total cell lysate incubated at 50°C after 2 h and 20 h. In parallel, a fraction of total cell lysate was incubated at 70°C and activity after 2 h and 20 h incubation was tested. More detailed expression tests of strains showing activity were then used to determine which recombinant vector gives the highest titer of protease.
In addition, processing conditions were optimized for four novel proteases and results are reported in the respective chapters (Table 2).

Alongside establishing the properties of novel proteases described in the upcoming chapters, initial characterization tests were performed for four additional proteases. Two proteases originating from *Thermus* (ZP_03496757.1 and ZP_03496880.1) were recombinantly produced in *E. coli* C43(DE3). The activity was solely found in the culture medium obtained after 24 h incubation at 37°C in the presence of 0.25% arabinose as inducer. In addition, recombinant expression was confirmed for two *Geobacillus thermodenitrificans* NG80-2 proteases YP_001125483.1 (NG483Sbt) and YP_001124317.1 (NG317Sbt). Expression of NG483Sbt subtilase required 24 h incubation at 30°C after induction with arabinose (0.25%) in the late log-phase. For isolation, a heating step was used (15 h 65°C). The re-suspended insoluble fraction showed activity in skim-milk plate assay after the heating step. On the other hand, optimal expression conditions for NG317Sbt subtilase included 48 h incubation at 17°C after induction with 0.1 mM IPTG in the late log-phase. This protease is located in the cytoplasm and in the total membrane fraction. By applying the heating step (50°C for 2 h) to the re-suspended insoluble fraction and cell-free extract followed by centrifuge step resulted in enriched protease activity in the respective supernatants. Functional expression of several bacterial subtilases (Table 2) was not observed in liquid cultures in spite of confirmed activity on plates. Further broader expression studies (autoinduction medium, addition of metal ions) should help.Since activity on plates does not always correspond to expression of active

subtilases in liquid cultures, probably due to specific growth and processing conditions, we tested the expression in liquid culture for few putative subtilases that were inactive in the plate assay. Two archaeal subtilases NP_579399.1 from *Pyrococcus furiosus* and YP_931297.1 *Pyrobaculum islandicum* were tested this way. Regrettably, we could not functionally express any of the putative archaeal proteases in our expression systems. Further immunoblotting assays can indicate whether these enzymes are expressed at all. A phylogenetic classification divides peptidases in terms of families and clans. For example, SB represents the subtilisin clan of serine peptidases and contains two families: S8 (subtilisin-like protease family) and S53 (sedolisin family). Siezen *et al.* reported[33] over 200 subtilisin-like proteases he termed subtilases. They were found in all kingdoms of life. By identifiying the structural core elements, the authors suggest further classification. The subtilase superfamily enzymes can be grouped into six families: proteinase K, thermitase, subtilisin, pyrolysin, kexin, and lantibiotic peptidase. The majority of active extremophilic proteases selected by us belong to the thermitase family, followed by proteinase K group and pyrolysin family, and subtilisin family (Figure 1).

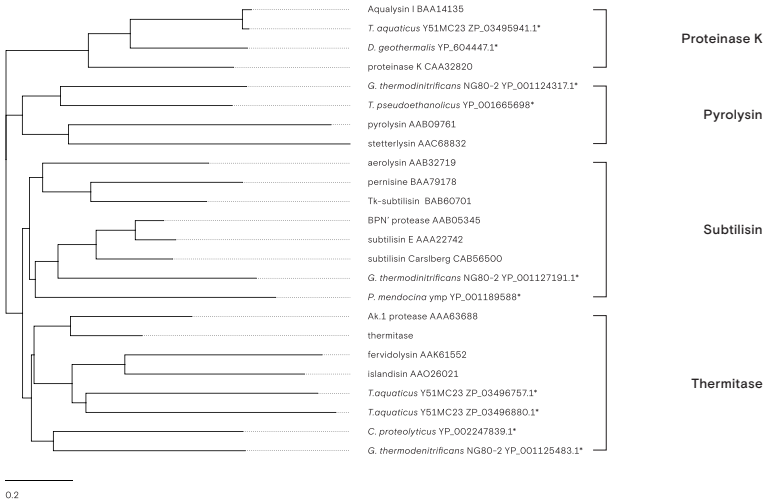



**FIGURE 1.**

Phylogenetic tree based on sequence alignment of literature described subtilases and functionally expressed subtilases of this study (indicated with an asterisk). The figure was generated with Geneious v5.6 software.





Members of the subtilisin group are only found in microorganisms and include mostly enzymes from Bacillus, where the subgroup of true subtilisins is best known[33]. Enzymes of this subgroup lack disulfide bonds and have two characteristic $Ca^{2+}$ binding sites called Ca1 (strong binding) and Ca2 (weak)[34]. Only few described members of the subtilisin group (aerolysin, pernisine and Tk-subtilisin) originate from hyperthermophilic organisms. The subtilisin-like proteases are expressed as precursors containing an N-terminal presequence, an N-terminal prosequence and a mature catalytic domain. Long C-terminal extensions are rare, but were predicted in some subtilases studied here by inspection of sequence alignments (Chapter 6 on *Pseudomonas subtilase*, *Pm*Sbt).

Thermitase is an extracellular, thermostable serine protease from the thermophilic microorganism *Thermoactinomyces vulgaris* and its family includes other subtilases originating from thermophilic organisms. The thermostability and the structure of thermitase were studied in detail[26,35]. There are two strong calcium binding sites and one weak calcium site. In the following chapters, four subtilases that cluster in the thermitase family were functionally expressed in *E. coli* (Figure 1). The biochemical characterization of a subtilase from *Coprothermobacter proteolyticus* is described in Chapter 4.

The members of the pyrolysin family are characterized by low sequence identity to subtilisins and carry sequence insertions and a long C-terminal domain. Two putative thermostable subtilases are clustered in this group, namely NG317Sbt (YP_001124317.1) and subtilase from *Thermoanaerobacter pseudoethanolicus* (YP_001665698).

Proteinase K is alkaline protease secreted by the mold *Engyodontium album*, and is similar to thermitase which is also heat stable and contains a cysteine close to the active site[36]. Proteinase K has two disulfide bonds, and like subtilisins it binds two calcium ions, albeit in different topology. Aqualysin I is another well-studied member of the proteinase K family whose crystal structure is solved[36]. Two extremophilic subtilases called *Taq*Sbt and *Dg*Sbt are clustered in this family and were identified and characterized in detail (Chapter 5).

## Conclusions

In our expression experiments two bottlenecks occurred. First, cloning of GC-rich sequences was successful for only 46% of the genes. For future projects, gene synthesis can provide a straightforward solution.

Another bottleneck is functional expression in recombinant host in case processing steps and secretion are needed to obtain active protease[19]. The full preprosequence of 24 subtilases was cloned into pBAD or pET-based vectors. From these 24 cloned genes, nine active subtilases were detected by plate screening and one enzyme was found in small-scale liquid culture screening. Furthermore, although promising as hyperthermophilic proteases, none of the archaeal subtilases could be obtained in the active recombinant form from transformed *E. coli* cells.

Another observation is that overexpression in *E. coli* was modest in all cases, with maximal levels of 5-50 mg per L of culture medium. At the moment, it is impossible to predict from the protein sequence whether a putative subtilase will be expressed and active in a selected host if close homologs with confirmed activity are unknown. The use of different recombinant system may also boost expression, as in the example of aqualysin I, where a recombinant expression level of 1 g/L using *Pichia* pastoris is reported[37]. The fact that subtilases of this research originate from extremophiles and require processing steps to become active may present an additional obstacle in obtaining new extremophilic subtilases using *E. coli* as a host.

Despite these limitations, in total ten putative extremophilic subtilases were functionally expressed using *E. coli* and found to be active. In the following chapters, the expression and properties of are described.

**Acknowledgments**

This work was supported by the program Integration of Biosynthesis and Organic Synthesis (IBOS-2; project 053.63.014), funded by The Netherlands Organization for Scientific Research (NWO). The authors would like to thank Linda Meekels for her work on the subtilases from Archaea.

**Author contribution**

AT performed the work, AT and DBJ wrote the chapter.





## References

01 — Adams, M., Perler, F. & Kelly, R. M. Extremozymes:expanding the limits of biocatalysis. *Nat. Biotechnol.* **13**, 662–668 (1995).

02 — Antranikian, G., Vorgias, C. & Bertoldo, C. Extreme environments as a resource for microorganisms and novel biocatalysts. *Adv. Biochem. Eng. Biotechnol.* **96**, 219–62 (2005).

03 — Gomes, J. & Steiner, W. The biocatalytic potential of extremophiles and extremozymes. *Food Technol. Biotechnol.* **42**, 223–235 (2004).

04 — Cowan, D. A. Thermophilic proteins: Stability and function in aqueous and organic solvents. *Comparative Biochemistry and Physiology - A Physiology* **118**, 429–438 (1997).

05 — Daniel, R. M., Cowan, D. A., Morgan, H. W. & Curran, M. P. A correlation between protein thermostability and resistance to proteolysis. *Biochem. J.* **207**, 641–644 (1982).

06 — Egorova, K. & Antranikian, G. Industrial relevance of thermophilic Archaea. *Curr. Opin. Microbiol.* **8**, 649–55 (2005).

07 — Van den Burg, B. Extremophiles as a source for novel enzymes. *Curr. Opin. Microbiol.* **6**, 213–218 (2003).

08 — Peek, K. Veitch, P. D., Prescott, M., Daniel, R.M., MacIver, B., & Bergquist, P. L. Some characteristics of a proteinase from a thermophilic *Bacillus* sp . expressed in *Escherichia col* : Comparison with the native enzyme and its processing in *E. coli* and *in vitro*. *Appl. Environ. Microbiol.* **59**, 1168–1175 (1993).

09 — Kirk, O., Borchert, T. V. & Fuglsang, C. C. Industrial enzyme applications. *Current Opinion in Biotechnology* **13**, 345–351 (2002).

10 — Gupta, R., Beg, Q. K. & Lorenz, P. Bacterial alkaline proteases: molecular approaches and industrial applications. *Appl. Microbiol. Biotechnol.* **59**, 15–32 (2002).

11 — Hartmann, R. & Meisel, H. Food-derived peptides with biological activity: from research to food applications. *Curr. Opin. Biotechnol.* **18**, 163–9 (2007).

12 — Namjoshi, S., Cacceta, R. & Benson, H. A. E. Skin peptides: Biological activity and therapeutic opportunities. *J. Pharm. Sci.* **97**, 2524–2542 (2008).

13 — Rao, M. B., Tanksale, A. M., Ghatge, M. S. & Deshpande, V. V. Molecular and biotechnological aspects of microbial proteases. *Microbiol. Mol. Biol. Rev.* **62**, 597–635 (1998).

14 — de Vos, W. M., Voorhorst, W. G. B., Dijkgraaf, M., Kluskens, L. D., van der Oost, J. & Siezen, R. J. Purification, characterization, and molecular modeling of pyrolysin and other extracellular thermostable serine proteases from hyperthermophilic microorganisms. *Methods Enzymol.* **330**, 383–393 (2001).

15 — Voorhorst, W. G. B., Warner, A., Vos, W. M. De & Siezen, R. J. Homology modelling of two subtilisin-like serine proteases from the hyperthermophilic archaea *Pyrococcus furiosus* and *Thermococcus stetteri. Protein Eng.* **10**, 905–914 (1997).

16 — Catara, G., Ruggiero, G., La Cara, F., Digilio, F. A., Capasso, A. & Rossi, M. A novel extracellular subtilisin-like protease from the hyperthermophile *Aeropyrum pernix* K1: Biochemical properties, cloning, and expression. *Extremophiles* **7**, 391–399 (2003).

17 — Foophow, T. Tanaka, S. I., Angkawidjaja, C., Koga, Y., Takano, K. & Kanaya, S. Crystal structure of a subtilisin homologue, Tk-SP, from *Thermococcus kodakaraensis*: Requirement of a C-terminal β-jelly roll domain for hyperstability. *J. Mol. Biol.* **400**, 865–877 (2010).

18 — Foophow, T., Tanaka, S., Koga, Y., Takano, K. & Kanaya, S. Subtilisin-like serine protease from hyperthermophilic archaeon Thermococcus kodakaraensis with N- and C-terminal propeptides. *Protein Eng. Des. Sel.* **23**, 347–55 (2010).

19 — Kannan, Y., Koga, Y., Inoue, Y., Haruki, M., Takagi, M., Imanaka, T. & Morikawa, M. Active subtilisin-like protease from a hyperthermophilic Archaeon in a form with a putative prosequence. *Appl. Environ. Microbiol.* (2001).

20 — Kluskens, L. D. Voorhorst, W. G. B., Siezen, R. J., Schwerdtfeger, R. M., Antranikian, G., van der Oost, J. & de Vos, W. M. Molecular characterization of fervidolysin, a subtilisin-like serine protease from the thermophilic bacterium *Fervidobacterium pennivorans. Extremophiles* **6**, 185–94 (2002).

21 — Toogood, H. S., Smith, C. A., Baker, E. N. & Daniel, R. M. Purification and characterization of Ak.1 protease, a thermostable subtilisin with a disulphide bond in the substrate-binding cleft. *J. Biochem.* **328**, 321–328 (2001).

22 — Munro, G. K., McHale, R. H., Saul, D. J., Reeves, R. a & Bergquist, P. L. A gene encoding a thermophilic alkaline serine proteinase from *Thermus* sp. strain Rt41A and its expression in *Escherichia coli. Microbiology* **141**, 1731–8 (1995).

23 — Peek, K., Daniel, R. M., Monk, C., Parker, L. & Coolbear, T. Purification and characterization of a thermostable proteinase isolated from *Thermus* sp. strain Rt41A. *Eur. J. Biochem.* **207**, 1035–44 (1992).

24 — Matsuzawa, H., Hamaoki, M. & Ohta, T. Production of thermophilic extracellular proteases (aqualysins I and II) by *Thermus aquaticus* YT-1, an extreme thermophile. *Agric. Biol. Chem.* **47**, 25–28 (1983).

25 — Gödde, C., Sahm, K., Brouns, S. J. J., Kluskens, L. D., van der Oost, J., de Vos, W. M. & Antranikian, G. Cloning and expression of islandisin, a new thermostable subtilisin from *Fervidobacterium islandicum*, in *Escherichia coli. Appl. Environ. Microbiol.* **71**, 3951–3958 (2005).

26 — Teplyakov, A. V, Kuranova, I. P., Harutyunyan, E. H. & Vainshtein, B. K. Crystal structure of thermitase at 1.4 A resolution. *J. Mol. Biol.* **214**, 261–279 (1990).

27 — Schiraldi, C. & De Rosa, M. The production of biocatalysts and biomolecules from extremophiles. *Trends Biotechnol.* **20**, 515–21 (2002).

28 — Bryan, P. N. Prodomains and protein folding catalysis. *Chem. Rev.* **102**, 4805–16 (2002).

29 — Ikemura, H. & Inouyes, M. Requirement of Pro-sequence for the production ofactive subtilisin E in *Escherichia coli. J. Biol. Chem.* **262**, 7859–7864 (1987).

30 — Lee, Y. C., Koike, H., Taguchi, H., Ohta, T. & Matsuzawa, H. Requirement of a COOH-terminal pro-sequence for the extracellular secretion of aqualysin I (a thermophilic subtilisin-type protease) in *Thermus thermophilus. FEMS Microbiol. Lett.* **120**, 69–74 (1994).

31 — Kim, D. W. & Matsuzawa, H. Requirement for the COOH-terminal pro-sequence in the translocation of aqualysin I across the cytoplasmic membrane in *Escherichia coli. Biochem. Biophys. Res. Commun.* **277**, 216–20 (2000).

32 — Terada, I., Kwon, S. T., Miyata, Y., Matsuzawa, H. & Ohta, T. Unique precursor structure of an extracellular protease, aqualysin I, with $NH_2$- and COOH-terminal pro-sequences and its processing in *Escherichia coli. J. Biol. Chem.* **265**, 6576–81 (1990).

33 — Siezen, R. J. & Leunissen, J. A. Subtilases: the superfamily of subtilisin-like serine proteases. *Protein Sci.* **6**, 501–23 (1997).

34 — Kristjansson, M. M. in *Thermostable proteins- Structural stability and design* (eds. Sen S. & Nilsson, L.) 67–104 (CRC Press, Taylor &Francis Group, Boca Raton, Florida, USA, 2011).

35 — Kleine, R. Properties of thermitase, a thermostable serine protease from *Thermoactinomyces vulgaris. Acta Biol. Med. Ger.* **41**, 89–102 (1982).

36 — Rawlings, N. D. & Salvesen, G. *Handbook of proteolytic enzymes.* (Academic Press, Elsevier, 2013).

37 — Olędzka, G., Dąbrowski, S. & Kur, J. High-level expression, secretion, and purification of the thermostable aqualysin I from *Thermus aquaticus* YT-1 in *Pichia pastoris. Protein Expr. Purif.* **29**, 223–229 (2003).

CHAPTER — 04

# Proteolysin, a novel highly thermostable and cosolvent-compatible protease from the thermophilic bacterium *Coprothermobacter proteolyticus*

Ana Toplak[a], Bian Wu[a], Fabrizia Fusetti[a], Peter J.L.M. Quaedflieg[b], Dick B. Janssen[a]

[a] Biochemical Laboratory, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, 9747 AG Groningen, the Netherlands
[b] DSM Innovative Synthesis, 6167 RD Geleen, the Netherlands

Part of this chapter has been published in: Toplak A, Wu B, Fusetti F, Quaedflieg PJML, Janssen DB. Proteolysin, a novel highly thermostable and cosolvent-compatible protease from the thermophilic bacterium *Coprothermobacter proteolyticus*. Appl Environm Microbiol (2013) 79: 5625-5632.

## Abstract

Through genome mining of extremophiles we identified a gene encoding a putative serine protease of the thermitase subgroup of subtilases (E.C. 3.4.21.66) in the thermophilic bacterium *Coprothermobacter proteolyticus*. The gene was functionally expressed in *Escherichia coli* and the enzyme which we called proteolysin, was purified to near homogeneity from crude cell lysate by a single heat-treatment step. Proteolysin has a broad pH tolerance and is active at temperatures of up to 80°C. In addition, the enzyme shows good activity and stability in the presence of organic solvents, detergents and dithiothreitol, and remains active in 6 M guanidinium hydrochloride. Based on its stability and an activity profile, proteolysin can be an excellent candidate for applications where resistance to harsh process conditions is required.



## Introduction

Extreme environments in terms of extreme temperature, pressure, salinity and pH represent valuable resources of unique genes and biocatalysts. These extreme habitats detrimental to the most mesophiles support the life of unusual organisms (extremophiles). With increased thermal stability, thermozymes (enzymes from hyper- and thermophiles) are valuable in developing novel bioindustrial routes[1,2]. In the search for novel thermostable endoproteases, we explored available genome sequences from thermophiles, hyperthermophiles and some bioremediation organisms for homologs of subtilisin E. We proceed with cloning and screening of putative subtilases (Chapter 3) and focus on an active subtilase from *Coprothermobacter proteolyticus*, a Gram-positive anaerobic extreme thermophile. This organism exhibits proteolytic activity and the organism was proposed as a candidate for thermophilic organic solid waste degradation, making characterization of its proteases of special interest[3].

Here, we describe the heterologous expression of this novel protease, its isolation by one-step heat purification, and its biochemical properties. The results show that this protease, which we termed proteolysin, is a highly thermostable protease and shows high resistance to harsh conditions.

## Materials and methods

**Chemicals and reagents —** Chemicals and reagents as well as cloning of proteolysin is described in Chapter 3. In addition, recombinant plasmid lacking sequence for the C-terminal hexahistidine tag (PrlAs) was made using the reverse primer 5'-CACACACGAAGCTT**TAA**TG¬GTGTCCAGTTTACTGCAGCATAC-3' (HindIII site under¬lined, stop codon in bold). A proteolysin mutant (PrlAd) harboring the substitutions Cys182Ala and Cys201Ala was obtained by site-directed mutagenesis using the QuikChange kit (Stratagene) and pBAD-MBP-PrlA as template.

**Expression, processing and isolation of proteolysin —** *E. coli* C43(DE3) cells containing an expression construct were grown overnight at 37°C in LB medium (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl) supplemented with 0.1% glucose and ampicillin (100 µg/ml). The overnight culture was used to inoculate (0.1% v/v) TB medium (12 g/L tryptone, 24 g/L yeast extract, 0.4% (v/v) glycerol, 72 mM $K_2HPO_4$, 17 mM $KH_2PO_4$) containing ampicillin (100 µg/ml) and cells were grown at 37°C. Proteolysin synthesis was induced at 17°C when the $OD_{600}$ reached 0.8 by adding 0.25% arabinose, followed by cultivation for 48 h at 17°C with shaking. The cells were harvested (4°C, 15 min at 6,000 x g), washed, suspended in buffer (50 mM Hepes·NaOH, pH 7.5, 10 mM $CaCl_2$) and disrupted by sonication. The total cell lysate was subjected to heat treatment at 80°C for 3 h and centrifuged (4°C, 45 min at 30,000 x g). The supernatant was concentrated with an Amicon YM30 ultrafiltration membrane (Millipore, Billerica, MA, USA), followed by buffer exchange to 50 mM Hepes·NaOH, pH 7.5, 10 mM $CaCl_2$. Protein concentrations were determined by the Bradford assay using bovine serum albumin as the standard. The mature protein was stored at - 80°C.

**SDS-PAGE and immunoblotting —** Samples of total cell lysate taken during proteolysin processing were analyzed on 12% SDS-PAGE gels. Proteins were transferred to a nitrocellulose membrane (Amersham Biosciences) using a semidry blotting apparatus (Bio-Rad, Hercules, CA, USA). Immunodetection was performed using antiserum against the hexahistidine tag, a secondary horseradish peroxidase-coupled antiserum and the ECL system from Amersham Biosciences according to the instructions of the manufacturer. Photographs were taken using a Fujifilm LAS-3000 imaging system.

**Enzyme assays —** Proteolytic activities of cell extracts and purified enzyme samples were monitored on 5% skim milk LB agar plates. Samples were added to wells in the agar and proteolytic activity was scored after overnight incubation at 37°C by inspection for clearing zones. For further analysis, an azocasein assay[4] and an amidolytic assay[5] were employed. For the first, an azocasein solution (1.25% (w/v) in Hepes·NaOH buffer (100 mM, pH 7.5) containing $CaCl_2$ (1 mM)) was used as the substrate. After 20 min incubation with enzyme at a selected temperature, trichloroacetic acid was added to quench the reaction (4%, v/v). The mixture was cooled on ice for 10 min and spun down in a tabletop centrifuge (4°C, 10 min, 14, 000 x g). Next, NaOH was added to the supernatant to 0.4 M final concentration and the absorbance was measured at 440 nm. Reactions were performed in triplicate. One unit of activity (U) is defined as the amount causing a change of one absorbance unit (AU) per min. In the amidolytic activity assay 1 mM *N*-succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide (*N* suc-AAPF-*p*NP) was used as the substrate, added from a 20 mM stock solution in DMSO. Reactions were performed in Hepes·NaOH buffer (100 mM, pH 7.5) containing $CaCl_2$ (1 mM) and DMSO (10%, v/v) at 40°C and 70°C. The release of *p*-nitroaniline was monitored by spectrophotometry at 410 nm (ε= 9.4 $mM^{-1} \cdot cm^{-1}$ (at 40°C), ε= 11.5 $mM^{-1} \cdot cm^{-1}$ (at 70°C). One unit is defined as the amount of activity that produces 1 µmol *p*-nitroaniline per min. For determination of kinetic parameters, the substrate concentration was varied from 0.03 - 3 mM.

The pH optimum and cosolvent tolerance were determined with the amidolytic assay, using Britton-Robinson universal buffer (0.04 M acetic acid, 0.04 M $H_3BO_3$, 0.04 M $H_3PO_4$) in a pH range of 3.0 – 12.0 (NaOH)). The concentration of cosolvents (DMSO, DMF or ethanol) was varied from 10 to 60% (v/v). Enzyme activities were normalized to the activity in the presence of 10% of the respective cosolvent. The selectivity of proteases was measured with oxidized insulin B chain. The substrate was dissolved to 10 mg/ml in 0.05 mM $NH_4HCO_3$, pH 8.0. Substrate to enzyme ratios were set to 1200 or 2400 (*w/w*). The reaction was started by addition of enzyme, and continued at 37°C and 60°C. Samples were taken at 0 min, 5 min and 24 h, conversion was stopped by adding 5% TFA, and peptides were analyzed by LC-MS/MS to determine the cleavage preference of proteolysin and subtilisin A.

**Inhibitors, surfactants, denaturants and metal ions —** The effects of surfactants and denaturants were tested by adding these compounds to reaction mixtures. The serine protease inhibitor phenylmethylsulfonyl fluoride (PMSF), EDTA, and iodoacetate were tested at concentrations of 5 and 10 mM. Metal ions were

added as chloride salts of calcium, magnesium, manganese and nickel. Enzyme was preincubated with the respective reagents dissolved in 100 mM Hepes·NaOH, pH 7.5, for 1 h at 37°C, and residual activities were determined in duplicate with the amidolytic assay at 40°C where $CaCl_2$ was omitted from the assay buffer. The activity in the same buffer without potential inhibitor added was set at 100%.

**Thermostability —** The fluorescence-based thermal stability assay described by Ericsson *et al.*[6] was used to determine apparent melting temperatures of the proteins. Protein solutions (15 µl) in buffer (25 mM Hepes·NaOH, pH 7.5, 1 mM $CaCl_2$, 20% glycerol) were incubated with PMSF (2.5 µl of 100 mM stock in ethanol) for 10 min at room temperature. A solution of 7.5 µl of 100 x Sypro Orange (Molecular Probes, Life Technologies, San Diego, CA) dye was added to the samples in a thin walled 96-well PCR plate. The plates were sealed with Optical-Quality Sealing Tape and heated in an iCycler iQ Real Time PCR Detection System (BioRad, Hercules, CA, USA) from 20 to 99°C at a heating rate of 1.1°C/min. The wavelengths for excitation and emission were 490 and 575 nm, respectively. Fluorescence changes were monitored with a charge-coupled device (CCD) camera and the derivative of the fluorescence change against time was used to obtain the $T_{m,\,app}$. Enzyme thermostability was also determined by incubating the enzyme in Hepes·NaOH buffer (100 mM, pH 7.5) containing $CaCl_2$ (1 mM) at 70, 80, 90 and 100°C for different time periods. Thermostability in the presence of DTT (10 mM) at 70°C was also determined for pretreated enzyme with 10 mM DTT (1h incubation at 25°C). The residual activity was determined in duplicate with the amidolytic assay after cooling to 40°C.

**Sulfhydryl groups —** For quantitative determination of sulfhydryl groups in proteolysin, Ellman's test was used as described by Riener *et al.*[7].

**Peptide analysis by mass spectrometry —** To allow confirmation of the identity of isolated proteolysin, a protein band was excised from a 12% SDS gel and tryptic digestion was performed according to Saller *et al.*[8]. For LC-MS/MS analysis, tryptic peptides of proteolysin or oxidized insulin B chain digests were first diluted in 0.1% TFA, separated on capillary C18 LC column and analyzed on MALDI-TOF/TOF 4800 Proteomics Analyzer (Applied Biosystems, Foster City, CA, USA) as described by Saller *et al.*[8]. Peptide identification was carried out using the program Protein Pilot 4.0 (ABsciex, Foster City, CA, USA) searching against the UniProtKB/Swiss-Prot protein database, to which the sequences of





proteolysin, trypsin and keratin were added. In cleavage specificity experiments, the sequence of oxidized insulin B chain was added to the database. Matches to proteolysin- or insulin-derived peptides were accepted if they could be assigned with a confidence of identification probability higher than 99%.

For intact protein analysis, samples were mixed 1:1 (*v/v*) with matrix solution consisting of saturated with α-cyano-4-hydroxycinnamic acid (14 mg/ml) in a 3:1:2 (*v/v/v*) mixture of formic acid/water/2-propanol (FWI) (pH 1.3), and analyzed after spotting on a stainless-steel MALDI target. Mass spectra were recorded by MALDI-TOF on the Applied Biosystems proteomics analyzer mentioned above, which was operated in linear positive ionization mode. Bovine serum albumin was used for calibration.

**Enzymatic peptide synthesis —** For synthesis of the dipeptide Cbz-Phe-Phe-$NH_2$ an isopropanol-precipitated enzyme preparation (IPREP suspension (20% w/v) for details see Chapter 5) in anhydrous acetonitrile was mixed with N-terminally protected acyl donor Cbz-Phe-OCam (50 mM), the C-protected nucleophile H-Phe-$NH_2$ (1.5 equiv., 75 mM), and heat-activated 3Å crushed molecular sieves (200 mg/mL). The reaction mixture (0.1 mL) was shaken for 24 h at 400 rpm at 37°C. The reaction was quenched (0.5 mL of acetonitrile-water (3:1, v/v), 1% formic acid) and filtered, followed by analysis by HPLC-UV (detection at 220 nm, details are given in Chapter 5).

# Results

## Genome mining

To discover thermostable subtilisin-like enzymes (subtilases) that might resist harsh process conditions, the entire preprosequence of subtilisin E (EMBL AAA22742) was used as a query in Blast searches of sequences of thermophilic bacteria, Archaea and bioremediation organisms (BLASTp, Chapter 3).

We selected the protease from *Coprothermobacter proteolyticus* for further studies since the source organism exhibits high proteolytic activity[9] and grows at high temperature. The expressed protease has quite high sequence identity with subtilisin E (35%) and thermitase (46%). According to Siezen's classification of the subtilase superfamily[10], this enzyme, which we termed proteolysin (PrlA), clusters in the family of thermitase (Chapter 3, Figure 1).

Inspection of sequence alignments of proteolysin with previously characterized subtilases revealed that the catalytic residues are fully conserved, *i.e.* the catalytic triad (Asp51, His94, and Ser267) and the oxyanion hole (Asn187) (Figure 1). A structural model of proteolysin based on the crystal structure of thermitase (1THM) was constructed with Yasara[11]. The model showed that proteolysin can adopt the same fold as subtilisin E. The active site geometry appears to be conserved and two calcium sites (Ca1 and Ca2) present in thermitase are likely to be present in proteolysin as well. Additional surface loops in the model originate from two sequence insertions (Asn55-Lys65 and Asp231-Asp243). Two cysteines (Cys182 and Cys201) are positioned in the inner core of the proteolysin model and fit the criteria for disulfide bond formation[12,13]. These cysteines are not conserved in the most homologous subtilases.



```
                                        90        100       110       120       130
                                   ...|....|....|....|....|....|....|....|....|....|.
proteolysin                 82   LPQGKAADVVMNQLKNDPNVEYVEPNYIAHAFDVPNDTFFNPYQWNFYDY
thermitase                  1    --------------------------------YTPNDPYFSSRQYGPQK-
subtilisin E                75   AAAATLDEKAVKELKKDPSVAYVEEDHIAH-------EYAQSVPYGISQ-
subtilisin BPN'             76   AASATLNEKAVKELKKDPSVAYVEEDHVAH-------AYAQSVPYGVSQ-
subtilisin Carlsberg        74   AAKAKLDKEALKEVKNDPDVAYVEEDHVAH-------ALAQTVPYGIPL-
aqualysin  I                96   GFAAEMAPQALEAFRQSPDVEFIEADKVVR-----AWATQSPAPWGLDRI
proteinase K                74   GFAATLDENMVRVLRAHPDVEYIEQDAVVT-----INAAQTNAPWGLARI
Clustal Consensus                 1                                                :.

                                             140       150       160       170       180
                                        ...|....|....|....|.... |....|....|....|....|....|
proteolysin                 132  GMTSNGYVSNYGIQAVSAWNITK-GAGVKVAIIDTGVAYENYGAYTKAPD
thermitase                  17   ------------IQAPQAWDIAE-GSGAKIAIVDTGVQSN-------HPD
subtilisin E                116  ------------IKAPALHSQGYTGSNVKVAVIDSGIDSS-------HPD
subtilisin BPN'             117  ------------IKAPALHSQGYTGSNVKVAVIDSGIDSS-------HPD
subtilisin Carlsberg        115  ------------IKADKVQAQGFKGANVKVAVLDTGIQAS-------HPD
aqualysin  I                141  DQRD--------LPLSNSYTYTATGRGVNVYVIDTGIRTT-------HRE
proteinase K                119  SSTS--------PGTSTYYYDESAGQGSCVYVIDTGIEAS-------HPE
Clustal Consensus                                     * .  : ::*:*:           :

                                                  190       200       210       220
                                             ....|....|....|....|. ...|....|....|....|....|....
proteolysin                 181  LANTLFDTANAYDFVNNDTHA-NDDNSHGTHVAGTIAQSTNNGMGAAGIA
thermitase                  48   LAG---KVVGGWDFVDNDSTP-QNGNGHGTHCAGIAAAVTNNSTGIAGTA
subtilisin E                148  LN-----VRGGASFVPSETNPYQDGSSHGTHVAGTIAALNN-SIGVLGVS
subtilisin BPN'             149  LK-----VAGGASMVPSETNPFQDNNSHGTHVAGTVAALNN-SIGVLGVA
subtilisin Carlsberg        147  LN-----VVGGASFVAGEAYN-TDGNGHGTHVAGTVAALDN-TTGVLGVA
aqualysin  I                176  FGG------RARVGYDALGGNGQDCNGHGTHVAGTIGGVTY------GVA
proteinase K                154  FEG------RAQMVKTYYYSS-RDGNGHGTHCAGTVGSRTY------GVA
Clustal Consensus                      :        .          : ..**** **  .        * :

                                                  280       290       300           310
                                             ..|....|....|....|....|...    .|....|....|...
proteolysin                 273  SSGSTTLQNAIQYAYNKGVVIVCASGNDRRS----TVSYPAAYTQCIAVG
thermitase                  137  TVGNSGLQQAVNYAWNKGSVVVAAAGNAGNT----APNYPAYYSNAIAVA
subtilisin E                235  PTGSTALKTVVDKAVSSGIVVAAAAGNEGSSGSTSTVGYPAKYPSTIAVG
subtilisin BPN'             236  PSGSAALKAAVDKAVASGVVVVAAAGNEGTSGSSSTVGYPGKYPSVIAVG
subtilisin Carlsberg        233  PSGSTAMKQAVDNAYARGVVVVAAAGNSGSSGNTNTIGYPAKYDSVIAVG
aqualysin  I                258  -GVSTALDNAVKNSIAAGVVYAVAAGNDNAN---ACNYSPARVAEALTVG
proteinase K                240  -GYSSSVNSAAARLQSSGVMVAVAAGNNNAD---ARNYSPASEPSVCTVG
Clustal Consensus                         .: :. .      * : . *:**          *.   .  :*.

                                        320       330       340       350       360
                                        .|....|....|....|....|....|....|....|....|....|...
proteolysin                 319  STRFDGTRARYSNYGSALDIVAPGGDTSVDQNHDGYGDGILQQTFAEGSP
thermitase                  183  STDQNDNKSSFSTYGSVVDVAAPG--------------SWIYSTYPTST-
subtilisin E                285  AVNSSNQRASFSSAGSELDVMAPG--------------VSIQSTLPGGT-
subtilisin BPN'             286  AVDSSNQRASFSSVGPELDVMAPG--------------VSIQSTLPGNK-
subtilisin Carlsberg        283  AVDSNSNRASFSSVGAELEVMAPG--------------AGVYSTYPTST-
aqualysin  I                305  ATTSSDARASFSNYGSCVDLFAPG--------------ASIPSAWYTSDT
proteinase K                287  ASDRYDRRSSFSNYGSVLDIFGPG--------------TSILSTWIGG--
Clustal Consensus                      :    . :: :*. *. ::: .**             : .:   .
```

```
                          370        380       390        400       410
                           .|....|....|....|....|....|....|....|....|....|...
proteolysin            369 TDFAYYFFQGTSMASPHVAGVAALVLSAHPTYTNEQVRTALQSTAKDLGT
thermitase             217 ----YASLSGTSMATPHVAGVAGLLASQGRSASN--IRAAIENTADKISG
subtilisin  E          319 ----YGAYNGTSMATPHVAGAAALILSKHPTWTNAQVRDRLESTATYLGN
subtilisin  BPN'       320 ----YGAYNGTSMASPHVAGAAALILSKHPNWTNTQVRSSLENTTTKLGD
subtilisin  Carlsberg  317 ----YATLNGTSMASPHVAGAAALILSKHPNLSASQVRNRLSSTATYLGS
aqualysin  I           340 ---ATQTLNGTSMATPHVAGVAALYLEQNPSATPASVASAILNGATTGRL
proteinase  K          320 ---STRSISGTSMATPHVAGLAAYLMTLG-KTTAASACRYIADTANKGDL
Clustal Consensus                    .*****:***** *.       . :       : . :

                          420       430        440       450       460
                           .|....|....|....|....|....|....|....|....|....|...
proteolysin            419 AGWDKYYGYGLVNAYAAVNWTP----------------------------
thermitase             262 TG--TYWAKGRVNAYKAVQ------------------------------
subtilisin  E          366 SF---YYGKGLINVQAAAQ------------------------------
subtilisin  BPN'       367 SF---YYGKGLINVQAAAQ------------------------------
subtilisin  Carlsberg  364 SF---YYGKGLINVEAAAQ------------------------------
aqualysin  I           388 SG----IGSGSPNRLLYSLLSSGSGSTAPCTSCSYYTGSLSGPGDYNFQP
proteinase  K          367 SN----IPFGTVNLLAYN----------------------------NYQA
Clustal    Consensus                :                                 *       *
```

FIGURE 1.

Partial sequence alignment of proteolysin, thermitase, subtilisin Carlsberg, subtilisin E, subtilisin BPN', proteinase K and thermostabile subtilase aqualysin I. The numbering of amino acids corresponds to proteolysin sequence. The start of the mature protein is marked in red. Conserved residues, *i.e.*, the catalytic triad (Asp51, His94, and Ser267) and oxyanion hole Asn186 are marked in dark gray. Cysteines, if present, are yellow. Sequence insertions characteristic for proteolysin are highlighted in green.

## Production, processing and isolation of proteolysin using *E. coli*

For the production of proteolysin in *E. coli*, plasmid pBAD-MBP-PrlA was constructed. A DNA segment encoding maltose binding protein (MBP) was fused to the 5' end of the region encoding the preprosequence to improve solubility of the precursor, and sequence for a C-terminal hexahistidine tag was added to facilitate purification. The recombinant fusion protein was expressed in *E. coli* strain C43(DE3). No significant activity was detected in the culture fluid. After sonication of the collected cells, a total cell lysate (TCL) was obtained, which





contained little proteolytic activity. Proteolysin processing and maturation required a heat treatment step. The processing could be performed at 70°C for 12 h or at 80°C for 3 h. We observed some product already after 1 h of heat treatment at 80°C, but only after 3 h of heat treatment a single protein band of approximately 40 kDa was obtained on SDS-PAGE gels (Figure 2). For routine enzyme isolation, a heat treatment step of 180 min at 80°C was used. The yield of purified proteolysin from a 1 L culture was 20 mg. The enzyme had a specific activity of 4 U/mg in the amidolytic assay. The results show that proteolysin was not secreted and processed by the *E. coli* host, and that the precursor protein was autocatalytically converted to mature enzyme upon heat treatment.

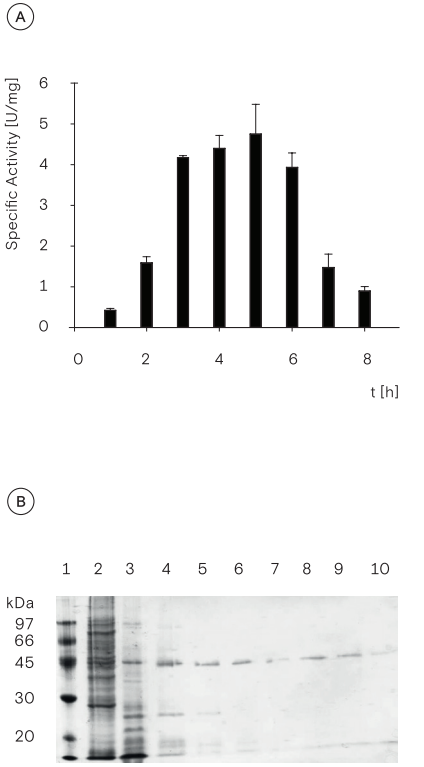



FIGURE 2.

Time course of proteolysin heat processing at 80°C. Panel A, Time-dependent heat activation at 80°C. Assays were done with *N*-suc-AAPF-*p*NA as the substrate. Panel B, electrophoretic analysis of proteolysin heat activation. TCA-precipitated samples were analyzed on a 12% SDS-polyacrylamide gel. Lanes: 1, molecular weight markers (GE Healthcare); 2, cell-free extract; 3-10, heat treatment at 80°C for 1h (lane 4), 2 h (lane 5), 3 h (lane 6), 4 h (lane 7), 5 h (lane 8), 7 h (lane 9) or 7 (lane 10).

The predicted amino acid sequence of the unprocessed recombinant proteolysin comprises 857 amino acids with a calculated total molecular mass of 92.8 kDa. The observed mass of the mature protein was 34,973 ± 3 Da, as determined by MALDI-TOF. To confirm the identity of the protein and to determine the N-terminus after maturation, tryptic digestion was performed and peptides were analyzed by MALDI-TOF/TOF. The peptide coverage was 67%, but no peptide segments were found in the Met1-Phe113 region. The molecular mass of the mature protein suggests that the sequence starts at Asp114 and ends with three additional amino acids (Leu328, Lys329, and Leu330) originating from the linker between the protein and the hexahistidine tag. Immunoblot analysis with antibodies against the hexahistidine tag confirmed its removal upon maturation of the protein (data not shown). Moreover, sequence alignment to thermitase supports Asp114 as the start of the mature protein.

To exclude an influence of the linker and hexahistidine tag on proteolysin processing, a gene encoding native proteolysin (termed PrIAs) including the stop codon was constructed. It was expressed and PrIAs was produced under the same conditions as above. The expression level and processing profile of the PrIAs, as well as its thermostability, were not impaired. The enzyme showed a similar activity as PrIA.

## Temperature profile and thermostability of proteolysin

Azocasein and *N*-succ-AAPF-*p*NA were used as the substrate for determination of temperature effects. Proteolysin hydrolyzed azocasein with optimal activity at 85°C, whereas subtilisin A was most active at 70°C under the conditions applied (Figure 3). The kinetic stability of proteolysin was investigated by following the loss of activity during incubation of the purified enzyme at high temperatures (70, 80, 90 and 100°C), using the amidolytic assay (Figure 4). Proteolysin appeared very stable at 70°C. Even after 20 h of incubation, the protein retained 35% of its initial activity towards *N*-suc-AAPF-*p*NA. In contrast, subtilisin A was completely inactivated after 1 h incubation at 70°C.





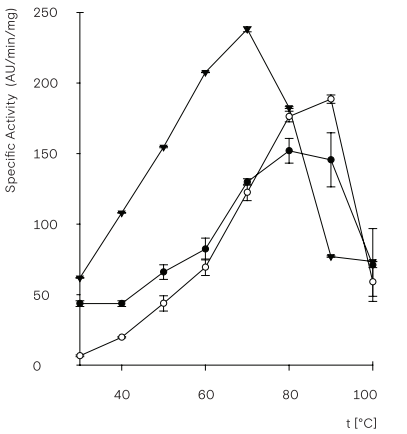



**FIGURE 3.**

Temperature profile of subtilisin A (▾), proteolysin PrIA (●), and proteolysin variant PrIAd (○). The apparent temperature optimum was determined in duplicate using the azocasein assay.

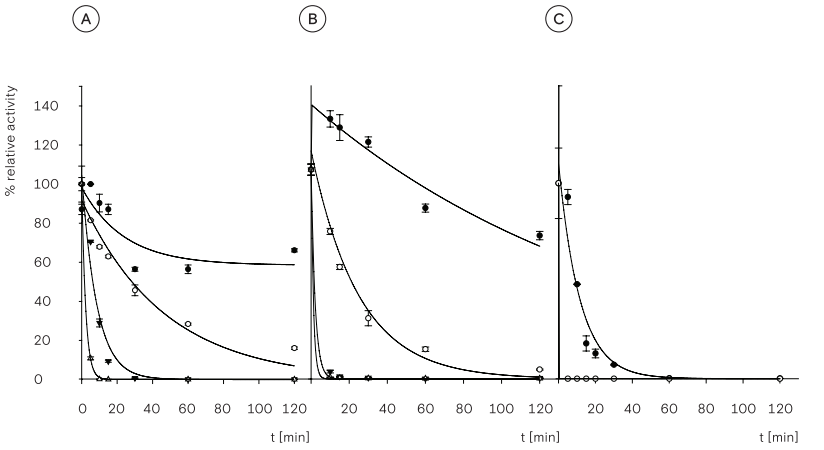



**FIGURE 4.**

Thermostability of proteolysin and subtilisin A. Enzymes were incubated at temperatures of 70°C (●), 80°C (○), 90°C (▾) or 100°C (Δ). The remaining activity at different times was determined with N-suc-AAPF-pNA as the substrate. Panel **A**, Proteolysin, PrIA; Panel **B**, Proteolysin, PrIAd; Panel **C**, subtilisin A.

The thermostability of proteolysin was also determined using the Sypro Orange assay. For this, enzyme activity was inhibited with PMSF. An apparent transition temperature ($T_{m,app}$) of 81 °C was observed, again indicating that the enzyme is much more thermostable than subtilisin A, for which a $T_{m,app}$ of 60 °C was measured.

A homology model for proteolysin was constructed with Yasara. The structural model suggested that a disulfide bond might be formed between Cys182 and Cys201, which are the only cysteines in the protein. Ellman's assay revealed the presence of 22 µmol of free sulfhydryls per mg of purified enzyme, in agreement with a sulfhydryl to enzyme ratio of 0.7. In a modified Ellman's test using cystamine[7], a sulfhydryl to enzyme ratio of 0.5 was found. These data suggest the occurrence of a disulfide bond or destruction of free sulfhydryls by a β-elimination reaction that may occur during heat treatment step[14]. To determine if the presence of a putative disulfide bond contributes to thermostability, a mutant carrying the substitutions Cys182Ala and Cys201Ala was constructed. The resulting variant (PrlAd) was expressed and produced under the conditions mentioned above, giving a four-fold lower protein yield as compared to proteolysin PrlA. PrlAd hydrolyzed azocasein with optimal apparent activity at around 90°C, slightly higher than what was found with the wild-type (Figure 2). The thermostability of variant PrlAd was only slightly affected by the mutations (Figure 4). Whereas the double mutant was more stable at low temperatures (70°C), the wild type showed slower inactivation at temperatures above 80°C. The presence of DTT (10 mM) reduced the half-life at 70°C by three-fold for both enzymes (PrlA and PrlAd, data not shown). This suggests that the Cys182-Cys201 pair does not contribute to the high thermostability of proteolysin.

## Catalytic properties

The catalytic properties of the novel protease were examined with *N*-suc-AAPF-*p*NP, a short peptide (oxidized insulin B chain), and azocasein (24.6 kDa), which is an azo-dye conjugate of casein, a large globular protein widely used in the food industry as a source of biologically active peptides[15]. The optimum pH of proteolysin was determined at 40°C and 70°C using *N*-suc-AAPF-*p*NP as the substrate (Figure 5). Proteolysin appeared active over a broad pH range (7.0-9.0) and showed maximal activity at approximately pH 8.0.





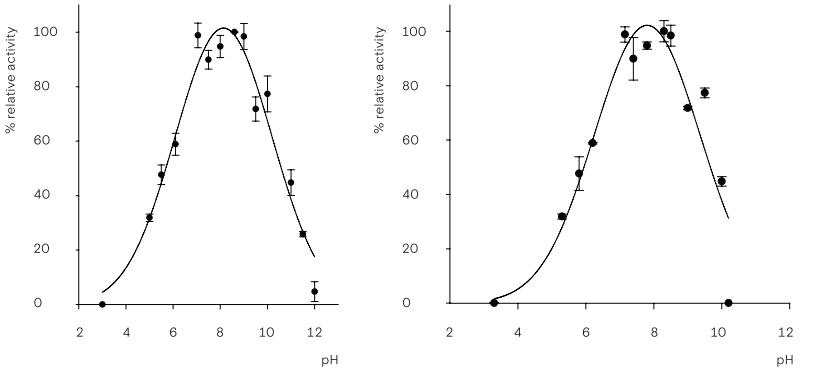



FIGURE 5.

pH profile of proteolysin PrlA. Britton-Robinson buffer was used in a pH range 3.0-12.0 (adjusted with NaOH). The reactions were performed in duplicate at 40°C (panel **A**) and 70°C (panel **B**). Data were fitted using SigmaPlot software.

The steady-state kinetic parameters of proteolysin PrlA, PrlAd and subtilisin A were determined with *N*-suc-AAPF-*p*NA and compared with those of other subtilases (Table 1). At 40°C proteolysin PrlA had a lower catalytic efficiency than its mesophilic counterparts. At elevated temperature, a four-fold increase of the $k_{cat}$ of proteolysin was accompanied by a similar increase in $K_m$, suggesting that the higher catalytic activity at elevated temperature is accompanied by less tight substrate binding. Interestingly, at 70°C the proteolysin variant lacking two cysteines (PrlAd) exhibited a catalytic efficiency of 335 $mM^{-1} \cdot s^{-1}$, which is 5-fold higher than that of the wild-type enzyme and similar to the $k_{cat}/K_m$ of subtilisin A with the same substrate at 40°C. Subtilisin A and the mutant proteolysin had similar activities at 70°C, but subtilisin was not very stable at this high temperature and its activity was lost after 1 h incubation (Figure 4).

| T | 40°C | | | 70°C | | | DTT (40°C) | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Enzyme** | $k_{cat}$ [$s^{-1}$] | $K_m$ [mM] | $k_{cat}/K_m$ [$mM^{-1}$ $s^{-1}$] | $k_{cat}$ [$s^{-1}$] | $K_m$ [mM] | $k_{cat}/K_m$ [$mM^{-1}$ $s^{-1}$] | $k_{cat}$ [$s^{-1}$] | $K_m$ [mM] | $k_{cat}/K_m$ [$mM^{-1}$ $s^{-1}$] | reference |
| Proteolysin PrlA | 25±1 | 0.41±0.07 | 62 | 110±5 | 1.58±0.13 | 69 | 37±1 | 0.57±0.04 | 65 | this work |
| Proteolysin PrlAd | 111±4 | 0.33±0.03 | 335 | 248±24 | 1.96±0.3 | 126 | 124±6 | 0.66±0.07 | 189 | this work |
| Subtilisin A | 551±62 | 1.1±0.3 | 489 | 539±70 | 1.06±0.3 | 508 | 40±3 | 1.57±0.25 | 23 | this work |
| Aqualysin I | 33 | 1.2 | 27.5 | ND | ND | ND | ND | ND | ND | [16] |
| Subtilisin BPN' | 480 | 0.29 | 1655 | ND | ND | ND | ND | ND | ND | [16] |

ND, not determined

TABLE 3.

Kinetic parameters of proteases measured with *N*-suc-AAPF-*p*NA as the substrate. Reaction temperatures were 40°C or 70°C. For assays in the presence of DTT, proteolysins were preincubated with 10 mM DTT at 25°C for 1 h, after which activities were determined. All incubations were performed in 100 mM Hepes·NaOH buffer, pH 7.5, containing 1 mM $CaCl_2$, 10% DMSO (*v/v*) and where indicated 10 mM DTT.

At 40°C, both proteolysin PrlA and PrlAd showed slightly higher $k_{cat}$ values in the presence of the reducing agent dithithreitol (DTT), but the increase in $K_m$ suggests weaker substrate binding (Table 1). High concentrations of DDT sometimes inactivate enzymes by metal chelation or by reducing essential disulfide bonds. Proteolysin and the mutant lacking two cysteines tolerated DTT concentrations up to 10 mM, indicating that in mature proteolysin the two cysteines do not form a DTT-reducible disulfide bond that is important for activity.





## Cleavage specificity

Oxidized insulin B chain was used as the substrate for determining the cleavage specificity of proteolysin. The reactions were performed at 37°C and the hydrolysis products were analyzed by LC-MS/MS. The primary cleavage site of proteolysin was found between Ala14 and Leu15 (Figure 6). Prolonged incubation (24 h) revealed additional cleavage sites after Phe1, Asn3, Gln4, Leu17, Cys19, Phe24, Phe25, Tyr25 and Lys29, which shows that proteolysin has a relaxed cleavage site specificity.

For subtilisin A, the primary cleavage site was between Leu15 and Tyr16 and prolonged incubation resulted in complete degradation of the substrate chain. The cleavage profile of proteolysin acting on insulin B chain did not change when the reaction temperature was raised to 60°C, whereas subtilisin A was inactivated after a prolonged incubation. When compared to other thermostable subtilases, proteolysin shares the primary cleavage site preference with pernisine[17] and shows a slight preference for hydrophobic amino acids residues at the P1 position, which is similar to other thermostable subtilases[18].

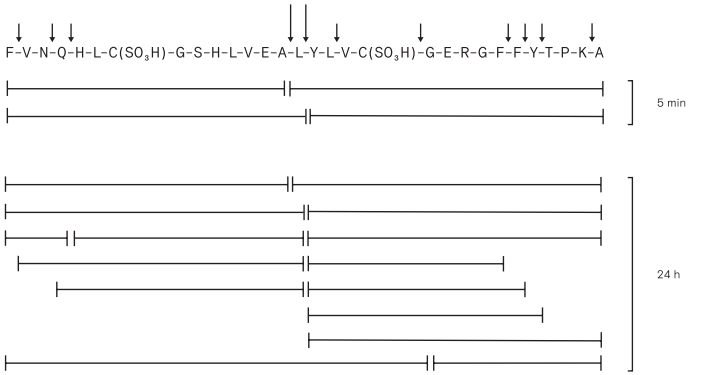



FIGURE 6.

Cleavage sites of proteolysin in oxidized insulin B chain. The diagram shows peptides identified by LC-MS/MS after 5 min and 24 h incubations with proteolysin at 37°C. Major and minor cleavage sites are denoted by large and small arrowheads, respectively.

## Metal ions and inhibitors

In most thermostable subtilases a bound calcium ion is essential for enzyme stability[19]. The number of calcium-binding sites varies from zero (Tk-SP) to seven (Tk-subtilisin)[20]. The structural models suggest that proteolysin may have two strong calcium-binding sites which are also present in thermitase[21]. Therefore, the effect of various cations on the amidolytic reaction was tested. The results showed that bivalent metal cations increased the activity of proteolysin (Figure 7), whereas the metal chelator EDTA (10 mM) decreased the activity to 87%. The serine protease inhibitor PMSF completely abolished the activity of proteolysin.

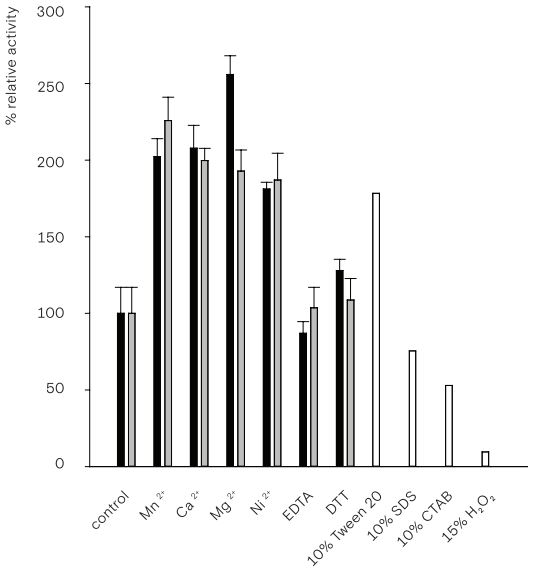



FIGURE 7.

The effect of different agents on proteolysin PrIA activity. The residual activity after 1 h incubation at 37°C with the compound indicated was measured with *N*-suc-AAPF-*p*NP as the substrate in duplicate. Activity with 10 mM additive is indicated with black bars, with 5 mM as grey bars. Activity with surfactants is indicated with white bars.





## Proteolysis under harsh conditions

The addition of cosolvents such as DMSO or DMF can be used to improve the solubility of peptides. Therefore, we tested the tolerance of proteolysin to DMSO, DMF and ethanol. Amidolytic assays were performed in the presence of various levels of cosolvents in the range of 10-60% (*v/v*). Proteolysin exhibits higher organic solvent tolerance than subtilisin A (Figure 8). However, even though subtilisin A is more susceptible to inactivation by cosolvents than proteolysin, the higher catalytic activity towards the model substrate compensates for this and subtilisin A still displays a higher catalytic activity in the presence of cosolvents.

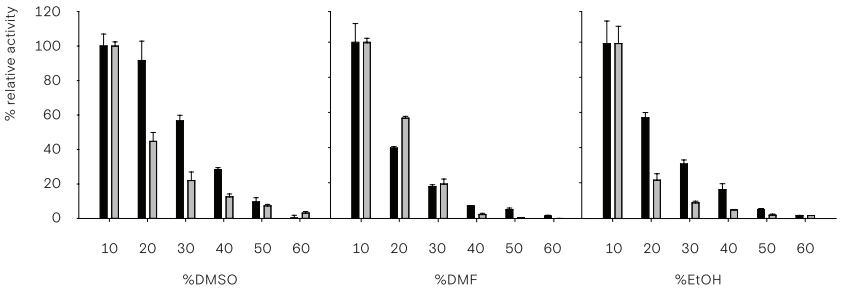



FIGURE 8.

Cosolvent tolerance. The effect of DMSO, DMF, and ethanol on the activity of proteolysin (black bars) and subtilisin A (grey bars). Activities were determined using the amidolytic assay at 40°C. Activities are normalized to the activity in the presence of 10% cosolvent.

Surface-active compounds or denaturing agents can be added to protein-containing substrates in order to increase substrate availability. We observed that proteolysin can tolerate up to 6 M guanidinium hydrochloride, which gave a 40% reduction of activity after 1 h incubation at 40°C. Subtilisin A was completely inactivated under the same conditions. The anionic surfactant SDS (10%, *w/v*) and the cationic surfactant CTAB (5%, *v/v*) reduced the activity of proteolysin, but

the non-ionic surfactant Tween 20 (10%, *v/v*) activated proteolysin up to 178%. Moreover, after 1 h of incubation in the presence of the oxidizing agent hydrogen peroxide (15% *v/v*), the residual activity dropped to 10% (Figure 7).

### Peptide synthesis

The possibility to use the proteolysin for biocatalytic peptide synthesis under anhydrous conditions was tested. A possible method for the preparation of dry enzyme is precipitation of the enzyme from buffer by an organic solvent i.e. isopropanol. Proteolysin, PrIA showed modest synthetic activity yielding 34% of dipeptide (Cbz-Phe-Phe amide) in neat acetonitrile after 24 h incubation at 37°C.

## Discussion

In this work, we describe the purification and biochemical characterization of proteolysin, one of the two putative extracellular proteases encoded in the genome of the extreme thermophile *C. proteolyticus*[22]. The organism was isolated from a thermophilic digester (55°C) fermenting tannery waste and cattle manure[23], and plays an important role in biogas production by releasing hydrogen that is used by methanogenic organisms[22].

Proteolysin belongs to the thermitase subgroup of the superfamily of subtilases. Produced as a precursor in *E. coli* cells, it requires processing for conversion to the mature form. The isolated yield of proteolysin produced in the mesophilic host *E. coli* was about 20 mg/L culture, which is sufficient for characterizing the enzyme and mutants thereof. The enzyme can be rapidly isolated by simultaneous processing and purification in a single heat-treatment step. Like many extracellular enzymes from thermophiles[24], proteolysin is active at temperatures higher than the optimum growth temperature of its *C. proteolyticus* host (63°C). Proteolysin can hydrolyze the large globular protein azocasein at temperatures as high as 90°C. The activity towards the synthetic substrate *N*-suc-AAPF-*p*NA is comparable to that of other thermophilic subtilases[25–27]. At elevated temperature (70–80°C), the enzyme is stable and displays a higher catalytic activity than at 40°C. By replacing both cysteines we obtained a variant with five-fold increased catalytic efficiency





at 40°C and with a similar thermostability and temperature/activity profile as the wild-type PrIA. When used in chemoenzymatic peptide synthesis in neat organic solvent, proteolysin PrIA was active, however the yield of dipeptide Cbz-Phe-Phe amide was poor as compared to other subtilases (Chapter 5).

Thermostability is often correlated with high tolerance to organic solvents[28]. Proteolysin indeed tolerated organic cosolvents in hydrolytic reactions much better than commercially available subtilisin A. Because of its thermostability, relaxed specificity and resistance to routinely used protein denaturants and DTT, proteolysin is a candidate for proteomics studies when protein digestion at extreme conditions is required. Heat pretreatment and the use of additives such as surfactants, organic solvents, and urea are commonly used to improve protein solubility and facilitate complete digestion, e.g. for peptide mapping[29,30]. In conclusion, the convenient production in E. *coli*, the high thermostability, the broad pH range, and its high tolerance to surfactants and cosolvents make proteolysin an attractive candidate for proteolysis under demanding reaction conditions.

**Acknowledgements**

We thank Dr. Hein J. Wijma for useful discussions and for making the model of proteolysin. This project is part of the Integration of Biosynthesis and Organic Synthesis program (IBOS-2; project 053.63.014) funded by The Netherlands Organisation for Scientific Research (NWO) and Advanced Chemical Technologies for Sustainability (ACTS). Peter J.L.M. Quaedflieg is employed by DSM. Some of DSM's products are manufactured with the use of proteases.

**Author contribution**

AT did experimental work with the proteases, FF performed mass spectroscopy analyses, BW, PJLMQ and DBJ supervised the work. AT, BW and DBJ wrote the paper.

# References


01 — Antranikian, G., Vorgias, C. & Bertoldo, C. Extreme environments as a resource for microorganisms and novel biocatalysts. *Adv. Biochem. Eng. Biotechnol.* **96**, 219–62 (2005).

02 — Gomes, J. & Steiner, W. The biocatalytic potential of extremophiles and extremozymes. *Food Technol. Biotechnol.* **42**, 223–235 (2004).

03 — Sasaki, K. , Morita, M., Sasaki D., Nagaoka, J., Matsumoto, N., Ohmura N. & Shinozaki, H. Syntrophic degradation of proteinaceous materials by the thermophilic strains *Coprothermobacter proteolyticus* and *Methanothermobacter thermautotrophicus. J. Biosci. Bioeng.* **112**, 469–72 (2011).

04 — Foophow, T. Tanaka, S. I., Angkawidjaja, C., Koga, Y., Takano, K. & Kanaya, S. Crystal structure of a subtilisin homologue , Tk-SP , from *Thermococcus kodakaraensi* : Requirement of a C-terminal β-Jelly roll domain for hyperstability. *J. Mol. Biol.* **400**, 865–877 (2010).

05 — Tanaka, T., Matsuzawa, H. & Ohta, T. Substrate specificity of aqualysin I, a bacterial thermophilic alkaline serine protease from *Thermus aquaticus* YT-1; Comparison with proteinase K, subtilisin BPN' and subtilisin Carlsberg. *Biosci. Biotechnol. Biochem.* **62**, 2161–2165 (1998).

06 — Ericsson, U. B., Hallberg, B. M., Detitta, G. T., Dekker, N. & Nordlund, P. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal. Biochem.* **357**, 289–98 (2006).

07 — Riener, C. K., Kada, G. & Gruber, H. J. Quick measurement of protein sulfhydryls with Ellman's reagent and with 4,4'-dithiodipyridine. *Anal. Bioanal. Chem.* **373**, 266–76 (2002).

08 — Saller, M. J., Fusetti, F. & Driessen, A. J. M. *Bacillus subtilis* SpoIIIJ and YqjG function in membrane protein biogenesis. *J. Bacteriol.* **191**, 6749–57 (2009).

09 — Klingeberg, M., Hashwa, F. & Antranikian, G. Properties of extremely thermostable proteases from anaerobic hyperthermophilic bacteria. *Appl. Microbiol. Biotechnol.* **34**, 715–719 (1991).

10 — Siezen, R. J. & Leunissen, J. A. Subtilases: the superfamily of subtilisin-like serine proteases. *Protein Sci.* **6**, 501–23 (1997).

11 — Krieger, E., Koraimann, G. & Vriend, G. Increasing the precision of comparative models with YASARA NOVA a self-parameterizing force field. *Proteins Struct. Funct. Genet.* **47**, 393–402 (2002).

12 — Pellequer, J.-L. & Chen, S. W. Multi-template approach to modeling engineered disulfide bonds. *Proteins* **65**, 192–202 (2006).







13 — Petersen, M. T., Jonson, P. H. & Petersen, S. B. Amino acid neighbours and detailed conformational analysis of cysteines in proteins. *Protein Eng.* **12**, 535–548 (1999).

14 — Volkin, D. B. & Klibanov, a M. Thermal destruction processes in proteins involving cystine residues. *J. Biol. Chem.* **262**, 2945–50 (1987).

15 — Silva, S. V. & Malcata, F. X. Caseins as source of bioactive peptides. *Int. Dairy J.* **15**, 1–15 (2005).

16 — Ikemura, H. & Inouyes, M. Requirement of Pro-sequence for the production ofactive subtilisin E in *Escherichia coli. J. Biol. Chem.* **262**, 7859–7864 (1987).

17 — Palmieri, G., Casbarra, A., Marino, G., Catara, G., Ruggiero, G., Capasso, A. & Rossi, M. High cleavage specificity of a subtilisin-like protease from a hyperthermophilic archaeon under extreme conditions. *Enzyme Microb. Technol.* **37**, 745–749 (2005).

18 — Mayr, J., Lupas, A., Kellermann, J., Eckerskorn, C., Baumeister, W. & Peters, J. A hyperthermostable protease of the subtilisin family bound to the surface layer of the archaeon *Staphylothermus marinus. Curr. Biol.* **6**, 739–49 (1996).

19 — Kristjansson, M. M. in *Thermostable Proteins- Struct. Stab. Des.* (Sen Lennart & Srikanta, N.) 67–104 (CRC Press 2011).

20 — Tanaka, S. I., Matsumura, H., Koga, Y., Takano, K. & Kanaya, S. Four new crystal structures of Tk-subtilisin in unautoprocessed, autoprocessed and mature forms: Insight into structural changes during maturation. *J. Mol. Biol.* **372**, 1055–1069 (2007).

21 — Gross, P. Calcium Binding to Thermitase. **266**, 2953–2961 (1991).

22 — Tandishabo, K., Nakamura, K., Umetsu, K. & Takamizawa, K. Distribution and role of *Coprothermobacter* spp . in anaerobic digesters. *J. Biosci. Bioeng.* **114**, 518–520 (2012).

23 — Ollivier, B. M., Mah, R. A., Ferguson, T. J., Boone, D. R., Garcia, J. L. & Robinson, R. Emendation of the genus Thermobacteroides: *Thermobacteroides proteolyticus* sp. nov., a proteolytic acetogen from a methanogenic enrichment. *Int. J. Syst. Bacteriol.* **35**, 425–428 (1985).

24 — Vieille, C. & Zeikus, G. J. Hyperthermophilic enzymes : Sources , uses , and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev.* **65**, 1–43 (2001).

25 — Foophow, T., Tanaka, S., Koga, Y., Takano, K. & Kanaya, S. Subtilisin-like serine protease from hyperthermophilic archaeon *Thermococcus kodakaraensis* with N- and C-terminal propeptides. *Protein Eng. Des. Sel.* **23**, 347–55 (2010).

26 — Peek, K., Daniel, R. M., Monk, C., Parker, L. & Coolbear, T. Purification and characterization of a thermostable proteinase isolated from *Thermus* sp. strain Rt41A. *Eur. J. Biochem.* **207**, 1035–44 (1992).

27 — Matsuzawa, H., Hamaoki, M. & Ohta, T. Production of thermophilic extracellular proteases (aqualysins I and II) by *Thermus aquaticus* YT-1, an extreme thermophile. *Agric. Biol. Chem.* **47**, 25–28 (1983).

28 — Liszka, M. J., Clark, M. E., Schneider, E. & Clark, D. S. Nature versus nurture: developing enzymes that function under extreme conditions. *Annu. Rev. Chem. Biomol. Eng.* **3**, 77–102 (2012).

29 — Riviere, L. R. & Tempst, P. in Curr. Protoc. *Protein Sci.* (John Wiley & Sons, Inc, 2001).

30 — Yu, Y. Q., Gilar, M., Lee, P. J., Bouvier, E. S. P. & Gebler, J. C. Enzyme-friendly, mass spectrometry-compatible surfactant for in-solution enzymatic digestion of proteins. *Anal. Chem.* **75**, 6023–6028 (2003).

CHAPTER — 05

# Peptide synthesis in neat organic solvents with novel thermostable proteases

Ana Toplak[a], Timo Nuijens[b], Peter J. L. M. Quaedflieg[b], Bian Wu[a], and Dick B. Janssen[a]

[a] Biochemical Laboratory, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, 9747 AG Groningen, the Netherlands
[b] DSM Innovative Synthesis, 6167 RD Geleen, the Netherlands

Part of this chapter has been published in:
Toplak A, Nuijens T, Quaedflieg PJML, Wu B, Janssen DB. Peptide synthesis in neat organic solvents with novel thermostable protease. Enz Microb Technol (2015) 73-74: 20-28.

Arif MI, Toplak A, Szymanski W, Feringa BL, Nuijens T, Quaedflieg PJML, Wu B, Janssen DB. One-step C -terminal deprotection and activation of peptides with peptide amidase from *Stenotrophomonas maltophilia* in neat organic solvent. Adv Synth Catal (2014) 10: 2197-2202

## Abstract

Biocatalytic peptide synthesis will benefit from enzymes that are active at low water levels in organic solvent compositions that allow good substrate and product solubility. To explore the use of proteases from thermophiles for peptide synthesis under such conditions, putative protease genes of the subtilase class were cloned from *Thermus aquaticus* and *Deinococcus geothermalis* and expressed in *E. coli*. The purified enzymes were highly thermostable and catalyzed efficient peptide bond synthesis at 80°C and 60°C in neat acetonitrile with excellent conversion (>90%). The enzymes tolerated high levels of *N,N*-dimethylformamide (DMF) as a cosolvent (40%–50% *v/v*), which improved substrate solubility and gave good conversion in 5+3 peptide condensation reactions. The results suggest that proteases from thermophiles can be used for peptide synthesis under harsh reaction conditions.



## Introduction

The discovery of novel peptide receptors and of bioactive peptides that modulate receptor- and enzyme activity have resulted in a growing importance of peptides for the pharmaceutical, food and cosmetics industries[1–4]. Synthesis of bioactive peptides can be performed chemo-enzymatically by enzymatic coupling of peptide segments obtained by solution- or solid-phase peptide synthesis (SPPS). For these conversions, a protease is applied in the synthetic direction[5]. In kinetically controlled coupling reactions, an activated acyl donor reacts with the enzyme and forms an acyl-enzyme intermediate that provides the N-terminal segment of the peptide product. If selectivity at the N-terminus is too restricted, substrate mimetics or strong leaving groups may be applied[6]. The intermediate undergoes an aminolytic reaction with the amine terminus of a nucleophilic peptide that becomes the C-terminal segment of the product. If these coupling reactions are carried out in an aqueous environment, the nucleophile needs outcompete water to cleave the acyl-enzyme intermediate and form a peptide bond; otherwise hydrolysis of the acyl-enzyme by water will occur. The influence of enzyme properties and reaction conditions on synthesis/hydrolysis ratios in aqueous systems is well explored for other applications of hydrolytic enzymes in coupling reactions[7,8]. Solely synthetic product without acyl donor hydrolysis should also be obtained when water is omitted from the reaction mixture, i.e. by using neat organic solvents[9,10]. This obviously requires stability of the enzyme during preparation and use of water-free protein[11].

Enzymatic coupling of peptide segments in organic solvents requires a broad-substrate-range enzyme that is stable and active in the absence of water. The industrial serine protease from *Bacillus licheniformis* (subtilisin A, subtilisin Carlsberg, Alcalase) has been used with success in peptide synthesis in neat organic solvent[12,13] with substrates carrying multiple side-chain protective groups to enhance solubility[14].

Peptide solubility may also be enhanced by increasing the reaction temperature and/or by the addition of an adequate cosolvent, such as dimethylformamide (DMF). Many enzymes show high thermal stability in anhydrous medium[15], but chemo-enzymatic peptide synthesis in neat organic solvent systems is generally performed at modest reaction temperatures, i.e. 25°C -37°C [10,14,16–19]. The use of higher temperatures is rare (50°C[20]). DMF can be added to increase substrate solubility in enzymatic coupling reactions[21,22], but it may negatively affect enzyme activity and, dependent on the substrate, the activity of subtilisin in neat DMF may be strongly reduced[23]. Protein engineering has been used to increase aqueous DMF resistance[24–26], but it remains to be established if such variants are active in peptide synthesis in anhydrous media mixed with DMF.

Enzymatic coupling at elevated temperatures in anhydrous medium with DMF as cosolvent would combine benefits of suppressed hydrolysis by the absence of water and increased substrate solubility due to high temperature and the presence of DMF. Since solvent stability and enzyme thermostability are often correlated[27,28], we decided to explore subtilisin-related enzymes from thermophilic organisms for peptide coupling under anhydrous conditions. We report the identification of two highly thermostable proteases and their synthetic performance in peptide synthesis at high temperature in acetonitrile-DMF solvent mixtures. In addition, we report the use of thermostable subtilases on preparative scale in economically favored peptide elongation in the *N→C* direction in a tandem reaction with an amidase.

## Materials and methods

**Chemicals and reagents —** Chemicals and reagents are listed in Chapter 3. Organic solvents were dried over 3 Å 4-8 mesh molecular sieves (Sigma-Aldrich, St. Louis, MO, USA), activated by heating at 200 °C. Peptide amides were obtained from Sigma-Aldrich and Bachem (Bubendorf, CH). C-terminal carbamoylmethyl (Cam, glycolamide) esters of peptides were synthesized by SPPS as reported by de Beer *et al.*[29].

**Enzyme production —** Cloning of the novel subtilases is described in Chapter 3. For the production of proteases the recombinant plasmids pET28a(+)-TAQ and pBADMycHisA(NdeI)-DG were introduced in *E. coli* C43(DE3) and transformants were grown at 37°C for 16 h on Luria Bertani (LB) agar plates containing an antibiotic (50 µg/mL kanamycin or 100 µg/mL ampicillin). Colonies were inoculated into 5 mL of LB broth containing 0.1% glucose and antibiotic. After 16 h of incubation at 37°C, 0.1 % (v/v) of the cultures were inoculated into 1 L of Terrific Broth (12 g/L tryptone, 24 g/L yeast extract, 0.4% (v/v) glycerol, 17 mM $KH_2PO_4$ and 72 mM $K_2HPO_4$) containing the appropriate antibiotic. For *Taq*Sbt protease production the medium was supplemented with 1% glucose and protease synthesis was induced in the late log phase with 0.3 mM IPTG, followed by continued growth at 37°C for 24 h with shaking at 200 rpm. In the case of *Dg*Sbt production, cultures were grown at 37°C until they reached the late log phase, cooled down and induced at 17°C with 0.25% L(+)-arabinose. Growth was continued for 48 h at 17°C with shaking at 200 rpm.

The protocol for protease processing and isolation was modified from the literature[30] to a one-step protocol. Cells were harvested (4°C, 15 min, 6,000 x g), suspended in 50 mM Hepes·NaOH, pH 7.5, containing 10 mM $CaCl_2$, and disrupted by sonication. For autocatalytic processing and purification the total cell lysate was subjected to heat treatment for 12 h at 65°C in the same Hepes buffer (Figure 2). During this step most proteins except for the protease are either degraded by proteolysis or precipitated by heat denaturation. After removal of denatured protein by centrifugation (4°C, 45 min, 30,000 x g), the supernatant was concentrated with an Amicon YM10 ultrafiltration membrane (Millipore, Billerica, MA, USA), followed by buffer exchange to 50 mM Hepes·NaOH, pH 7.5, containing 10 mM $CaCl_2$. Protein concentrations were determined by the Bradford assay using bovine serum albumin as a standard. The protocol yielded up to 25 mg of *Dg*Sbt protease or 50 mg of *Taq*Sbt protease from a 1 L culture. The processed proteins gave a single band of 28 kDa on an SDS-polyacrylamide gel (estimated purity >80%). The proteins were stored at -80°C.

Active subtilases were also isolated from the heat-treated centrifuged cell-free extract by precipitation with ammonium sulfate (70% sat. at 4°C, 30 min stirring at 4°C). The precipitate was collected by centrifugation (4°C, 15 min, 30 000 x g), washed with 80% acetone to remove remaining lipophilic compounds, re-suspended in 20 mM Hepes·NaOH, pH 7.5, containing 1 mM $CaCl_2$ and 20% glycerol, and used for anhydrous enzyme preparation as described below. If required, glycerol was removed by exchanging the buffer on a HiPrep 26/10 desalting column (GE Healthcare Bio-Sciences, Pittsburgh, PA, USA).

**Preparation of anhydrous enzyme for synthesis reactions —** Enzyme in 20 mM Hepes·NaOH, pH 7.5, containing 1 mM $CaCl_2$ and supplemented with sucrose (10 mg/mL) as lyoprotectant, was used for lyophilization. Anhydrous enzyme samples were prepared using several dried solvents in the same way as for isopropanol[31]. Briefly, to obtain isopropanol-precipitated and rinsed enzyme preparations (IPREP), organic solvent was added to the enzyme solution in buffer (20 mM Hepes·NaOH, pH 7.5, containing 1 mM $CaCl_2$ and 20% glycerol) to a ratio of 1:3, 1:6 or 1:9 (*v/v*, aq/solvent) until formation of a precipitate was observed. After vortexing, the precipitate was collected in a tabletop centrifuge (room temperature, 5,000 rpm, 5 min) and remaining liquid was removed. The pellet was twice suspended in the same organic solvent, followed by vortexing and centrifugation. Subsequent drying in a Speedvac at room temperature for 20 min gave a white powder (IPRED) that was active when redissolved in buffer. The dry enzyme was stable for at least one month when stored at room temperature (20-25°C). The IPRED preparations could be easily suspended in dry solvents, without a strong tendency to aggregate. Anhydrous preparations of subtilisin A were obtained by precipitation with *t*-BuOH as described by Chen *et al.*[32].

Active site titration with the irreversible serine protease inhibitor phenylmethanesulfonyl fluoride (PMSF) was used to determine the concentration of active enzyme present in enzyme samples. To 190 µL of PMSF solution of varying concentration (0-75 µM), prepared by mixing fresh 3 mM stock in ethanol and buffer (100 mM Hepes·NaOH, pH 7.5, with 1 mM $CaCl_2$), 10 µL of enzyme solution (approx. 0.4 mg/mL) was added. The mixture was vortexed and incubated for 10 min at room temperature. The remaining activity was then measured in the standard amidolytic assay as described below.

**Enzyme characterization —** Homology models of both subtilases were constructed with Yasara[33] based on the aqualysin I structure (PDB: 4DZT, 98% sequence identity) for TaqSbt and based on proteinase K structure (PDB: 2B6N, 59% sequence identity) for *Dg*Sbt.

General proteolytic activities were measured at several temperatures by following the release of soluble sulfanilamide-azopeptides from azocasein in a UV/VIS spectrophotometer at 440 nm, as described previously[34]. Amidolytic activity was quantified by following the release of *p*-nitroaniline from the standard proteolytic substrate *N*-succinyl-Ala-Ala-Pro-Phe *p*-nitroanilide (*N*-suc-AAPF-pNA) at 40°C in 100 mM Hepes·NaOH, pH 7.5, containing 1 mM $CaCl_2$, 10 % (v/v) DMSO and





substrate (0.3-4 mM)[34]. Residual activities of heated samples were assayed after cooling. For thermostability measurements, the thermofluor method[35] was applied as described in Chapter 4.

**MALDI / TOF / TOF analysis —** For mass spectrometry based protein identification, the samples of *Taq*Sbt and *Dg*Sbt were analyzed by SDS-PAGE. After Coomassie staining, bands corresponding to 28 kDa were excised, digested with trypsin as reported previously[34], and analyzed by LC-MS. Mass spectrometry was carried out with a MALDI-TOF/TOF 4800 Proteomics Analyzer (Applied Biosystems, Foster City, CA, USA) in the m/z range 600-4000. Peptide identification was done using the program ProteinPilot 4.0 (ABsciex, Foster City, CA, USA) against UniProtKB/Swiss-Prot protein sequence database[36] to which the sequences of trypsin and keratin were added. Trypsin specificity and default parameters were used for the search.

For intact protein analysis, samples were prepared using the ultra-thin layer method for MALDI[37] and mass spectra were recorded by MALDI-TOF on the Applied Bio-systems proteomics analyzer mentioned above, which was operated in linear positive ionization mode. Bovine serum albumin was used for calibration.

**Enzymatic peptide synthesis —** Tests to examine enzymatic synthesis of the dipeptide Cbz-Phe-Phe-$NH_2$ were carried out by mixing the IPRED suspension (10% *w/v*) in the desired anhydrous solvent or solvent/cosolvent mixture with the N-terminally protected acyl donor Cbz-Phe-OCam (50 mM) and the C-protected nucleophile H-Phe-$NH_2$ (1.5 equiv., 75 mM) in the same solvent and dried for 1 h over heat-activated 3Å crushed molecular sieves. Dry molecular sieves were also added to the reaction mixture (200 mg/mL, 0.1-0.2 mL), which was shaken for at least 24 h at 400 rpm at 60 or 80°C. Samples (20 µL) were taken at different times, quenched (0.5 mL of acetonitrile-water (3:1, *v/v*), 1% formic acid) and filtered before HPLC analysis.DMF tolerance was tested by performing enzymatic synthesis of Cbz-Phe-Phe-$NH_2$ as described above on 100 µL scale using IPRED enzyme preparations (10% *w/v*) containing 11 µM active *Taq*Sbt and 25 µM active *Dg*Sbt subtilases, 25 mg/mL 3 Å molecular sieves and various concentrations of DMF (0-100% *v/v*). The mixtures were incubated at 60°C for 1 h, quenched (100 µL DMSO and 0.5 mL acetonitrile-water (3:1, *v/v*), 1% formic acid solution), filtered and analyzed by HPLC-UV (220 nm, see below).

Reaction conditions for establishing the substrate scope of *Dg*Sbt and *Taq*Sbt in peptide synthesis were slight variations, as indicated in the text, of the following protocol for heptapeptide synthesis (Ac-Asp-Phe-Ser-Lys-Leu-Ala-Phe-$NH_2$). To the *Taq*Sbt or *Dg*Sbt IPRED preparation (110 mg), both 0.5 ml of water-free substrate solution mixture (20 mM acyl donor and 30 mM nucleophile dissolved in DMF: acetonitrile: $Et_3N$ (1:0.98:0.02)) and 10 mg 3 Å molecular sieves were added. The reaction mixture was incubated for 24 h at 60°C, and samples (20 µL) were taken, quenched with DMSO (80 µL), followed by a short centrifugation step to remove molecular sieves and enzyme. Conversions were estimated by HPLC-UV (220 nm, see below).

Water content in reaction mixtures was determined by Karl-Fischer titration (702SM Titrino, Metrohm). Water content typically was 0.12-0.13%, which mostly originated from enzyme, irrespective of the use of fresh or stored dried enzyme.

**Preparative scale bioconversions —** Kinetic coupling of Z-Gly-Tyr-OMe with H-Phe-$NH_2$ using *Dg*Sbt protease was performed as follows. To the *Stenotrophomonas maltophilia* amidase (PAM)[38] reaction mixture containing *N*-protected acyl donor Z-Gly-Tyr-OMe (8 mmol) in acetonitrile (1.5 ml), *C*-terminally protected Phe-$NH_2$ (10 equiv. 80 mmol) was directly added, as well as 90 mg IPREP (isopropanol rinsed enzyme precipitate) containing 4 mg of *Dg*Sbt and activated 3Å crushed molecular sieves (200 mg/ml). The reaction mixture was shaken at 400 rpm, 60°C. Samples were taken and quenched with DMSO (1:3, *v/v*). Reaction was stopped after 21 days.

Kinetic coupling of Z-Gly-Tyr-OMe with H-Phe-$NH_2$ using *Taq*Sbt protease was done as follows. In a parallel coupling reaction, to the purified *N*-protected acyl donor Z-Gly-Tyr-OMe (28 mmol) in acetonitrile (3 ml), *C*-terminally protected Phe-$NH_2$ (10 equiv. 280 mmol) was added, followed by 35 mg IPREP enzyme preparation (2.2 mg of *Taq*Sbt enzyme) and activated 3Å molecular sieves powder (100 mg/ml). The reaction mixture was shaken at 400 rpm at 60oC. Samples were taken and quenched with DMSO (1:3 *v/v*). After 6 days the reaction was stopped and analyzed. Conversions were estimated by HPLC and further identified by LC/MS.

**HPLC and LC/MS —** Peptide coupling reactions were analyzed by HPLC. The DMSO-formic acid-quenched samples (20 µL) were loaded on a Gemini-NX C18 5µm 110A column (250x4.60 mm, Phenomenex). The solvent system consisted of 0.01% formic acid in water (eluent A) and 0.08% formic acid in acetonitrile





(eluent B). The gradient program was: 0-13.5 min linear gradient from 10%-50% eluent B in A, continued for 6.5 min at 50% eluent B, then 10 min at 10% eluent B. Separation was carried out at 25°C, with detection at 220 nm. For dipeptide synthesis reactions, the response factors of starting material (Cbz-Phe-OCam), hydrolytic product (Cbz-Phe-OH) and synthetic product (Cbz-Phe-Phe-$NH_2$) were determined, and yields were calculated. In case of other synthetic reactions, where product was not available, the conversion was calculated based on the consumption of the acyl donor substrate and the formation of the hydrolytic product, where acyl donor and hydrolytic product were assumed to have similar response factors. The latter was confirmed by complete enzymatic hydrolysis of the model substrate Ac-Asp-Phe-Ser-Lys-Leu-OCam, which showed that the OCam group did not significantly contribute to absorbance at 220 nm.

The identity of synthetic products was confirmed on an LC/MS ion-trap system (LCQ Fleet, Thermo Scientific Waltham, MA, USA). The quenched reaction mixture or collected HPLC peak was loaded on an Alltech C18 3µ column (3 µm, 100x4.6 mm), with the same solvent system as for HPLC analysis. The elution gradient was as follows: 100% A for 2 min, 30 min linear to 40% B in A, followed by 5 min 40% B and 10 min 100% A, or alternatively applying 80% solvent B (for dipeptide analysis). Detection was at 220 nm and by ESI-MS operated in a positive ionization mode.

## Results and Discussion

### Enzyme selection and production

We previously reported results of genome mining to obtain novel thermostable endoprotease homologs of subtilisin E[34]. Among the six active subtilases that were reported, *Taq*Sbt from *Thermus aquaticus* Y51MC23 (subtilase ZP_03495941) and *Dg*Sbt originating from *Deinococcus geothermalis* (subtilase YP_604447) were functionally expressed in *E. coli*.
*Taq*Sbt has 98% sequence similarity to aqualysin I from *Thermus aquaticus* YT-139 and 59% to commercial subtilase Rt41A (PreTaq) from *Thermus* sp.[40], whereas identity of *Dg*Sbt to these protein sequences is 52% and 54%, respectively. Sequence alignment (Supplementary info) suggests that both enzymes are synthesized as a precursor consisting of an N-terminal signal sequence followed by a prodomain, the mature enzyme domain and an additional C-terminal

prodomain, which is characteristic for the aqualysin I-type subtilases[41]. The C-terminal prodomain used for translocation in the natural host can be omitted or will be removed autocatalytically during the heat treatment step through an intermolecular cleavage mechanism[42].

According to Siezen's classification of the subtilases[43], *Taq*Sbt and *Dg*Sbt are clustered in the proteinase K subgroup (Chapter 3, Figure 1). From sequence alignments (Figure 1) and homology models we observed that the overall subtilase fold and the topological position of the catalytic triad are conserved in *Taq*Sbt (numbering from the first amino acid of the ORF; residues Asp166, His197, Ser349), and in *Dg*Sbt (numbering from the first amino acid of the ORF; Asp171, His203, Ser355), as is the oxyanion hole (residues Thr348, Ser349 and Asn284 in TaqSbt; Thr354, Ser355 and Asn290 in DgSbt). Two disulfide bonds (Cys194-Cys226; Cys290-Cys321) with aqualysin I topology[44], as well as the two calcium binding sites of aqualysin I[45,46] are predicted for the *Taq*Sbt model. In the *Dg*Sbt model, disulfide bonds with identical topology are also predicted (Cys200-Cys232, Cys296-Cys327), but only one calcium binding site was found. Both proteases were successfully cloned and expressed in *E. coli* C43(DE3) using pBADMycHisA(NdeI) and pET28a+ vector systems.

Heat-shock activation and purification could be applied to *E. coli* cell-free extract to quickly isolate the thermostable enzymes (Supplementary info). Proteolytic activity was detected and was isolated from the heat-treated extracts of *E. coli* cells. The yield of pure *Taq*Sbt with the pET system reached 50 mg/L of culture, with a specific activity of 12 U/mg in the amidolytic assay, whereas the pBAD system for *Dg*Sbt production gave 25 mg of pure protein per L of culture with a specific amidolytic activity of 0.5 U/mg. Subtilase identity was confirmed by mass spectrometry of the whole proteins and by mass analysis of tryptic segments prepared by digesting proteins extracted from a polyacrylamide gel. The processing sites could be assigned by inspecting protein sequence alignments and from molecular weights of intact protein and of tryptic segments determined by MALDI/TOF/TOF. The protein sequence analysis and the observed molecular weight of 28, 2492 ± 3 Da for *Taq*Sbt correspond well to the calculated value of 28,2496 Da for the 296 amino acid long protein starting with Ala133 and ending with Ser428 (numbering from the first amino acid of the ORF), implying *Taq*Sbt is one Gly residue shorter than aqualysin I (MW calc. 28,3067 Da) at the C-terminus. In addition, the determined molecular weight of 28,163 ± 3 Da in combination with the results of protein sequence analysis suggest that Ala133 (numbering from the first amino acid of the ORF) is the first amino acid and Gly417 the last amino acid of the mature *Dg*Sbt.





## Biochemical and catalytic properties

The thermostability of both enzymes was investigated by measuring the loss of amidolytic activity during incubation of the purified enzyme at high temperatures (70, 80, 90 and 100°C) (Figure 1). Although *Dg*Sbt was stable at 70°C ($t_{1/2}$ = 2 h), *Taq*Sbt is clearly the more thermostable protease, with only 30% loss of activity after 1 h incubation at 90°C. The biphasic nature of the curves suggests some protease heterogeneity, which could be due to differences in degree of disulfide bond formation or metal binding. In the azocasein assay, *Dg*Sbt showed at temperature optimum of 69°C and *Taq*Sbt gave an optimum at 92°C (Figure 2). Thermostability was also measured by determining apparent melting temperatures ($T_{m,app}$) of PMSF-inhibited enzyme with the thermofluor method. The results showed that *Dg*Sbt had a $T_{m,app}$ of 72°C, whereas the $T_{m,app}$ of *Taq*Sbt could not be measured because the upper temperature limit of 99°C appeared too low for full *Taq*Sbt denaturation. The higher stabilities in the thermofluor assays are in agreement with autodigestion being inhibited by the presence of PMSF.

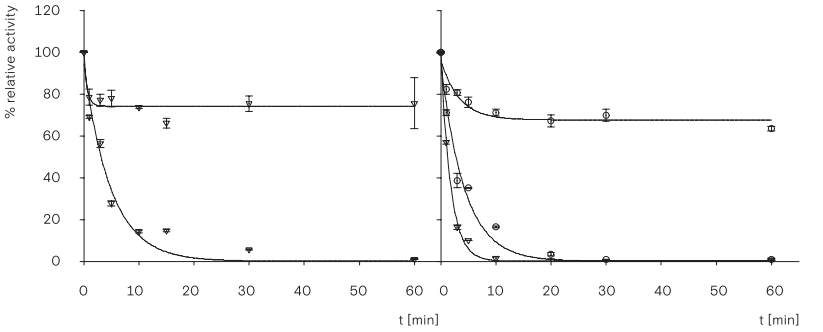



FIGURE 1.

Thermostability of TaqSbt (Panel A) and DgSbt (Panel B). DgSbt was incubated at temperatures of 70°C (●), 80°C (○), or 90°C (▾), whereas TaqSbt was incubated at 90°C (▾) and 100°C (Δ). The remaining activities at different times were determined after cooling to the standard assay temperature using *N*-suc-AAPF-*p*NA as the substrate. The reactions were performed in duplicate

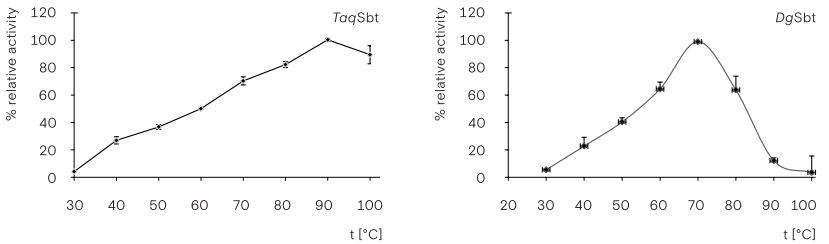



**FIGURE 2.**

Temperature profile of *Dg*Sbt and *Taq*Sbt. The apparent temperature optimum was determined in duplicate using the azocasein assay.

The optimum pH was determined at 40°C using *N*-suc-AAPF-*p*NP as the substrate. *Taq*Sbt appeared active over a broad pH range (6.5-9.5), whereas *Dg*Sbt showed maximal amidolytic activity at approximately pH 8.0 (data not shown). The steady-state kinetic parameters of *Taq*Sbt and *Dg*Sbt were also determined and compared with those of other subtilases (Table 1). The thermostable subtilases aqualysin I, *Taq*Sbt and *Dg*Sbt catalyzed hydrolysis of *N*-suc-AAPF-*p*NA with lower efficiency at 40°C than subtilisin A. Although the $k_{cat}$ values of thermostable proteases were in the same range, differences were found in substrate affinity, which was 2-6-fold higher in case of *Taq*Sbt than with the listed thermophilic enzymes (Table 1).

| Enzyme | $k_{cat}$ [$s^{-1}$] | $K_m$ [mM] | $k_{cat}/K_m$ [$mM^{-1}\ s^{-1}$] | reference |
|---|---|---|---|---|
| *Taq*Sbt | 25±3 | 0.62±0.16 | 47 | this work |
| DgSbt | 23±3 | 3.6±0.7 | 6.5 | this work |
| Subtilisin A | 551±62 | 1.1±0.3 | 489 | [34] |
| Aqualysin I | 33 | 1.2 | 27.5 | [46] |

**TABLE 1.**

Kinetic parameters of proteases. All reactions were performed in 100 mM Hepes·NaOH buffer, pH 7.5, containing 1 mM $CaCl_2$ and 10% DMSO (*v/v*) at 40°C using *N*-suc-AAPF-*p*NA as the substrate.





The effects of various metal ions, protease inhibitors and detergents on *Taq*Sbt and *Dg*Sbt activity were determined by measuring residual amidolytic activity after 1 h preincubation under appropriate conditions (Table 2). The presence of $Ca^{2+}$ slightly increased the activity of both subtilases. High concentrations of the metal chelator EDTA decreased the activity of *Taq*Sbt to 68% and of *Dg*Sbt to 40%, confirming the need of $Ca^{2+}$ ions for activity. The serine protease inhibitor PMSF abolished the activity of both proteases, confirming that both *Dg*Sbt and *Taq*Sbt are serine proteases.

| Salts, inhibitors, detergents | Concentration | % Residual activity[a] | |
|---|---|---|---|
| | | *Dg*Sbt | *Taq*Sbt |
| None | - | 100 | 100 |
| $Ca^{2+}$ | 10 mM | 120±13 | 128±1 |
| SDS | 10% (w/v) | 78±3 | 98±3 |
| CTAC | 12.5% (w/v) | 2.5±0.1 | 6.4±0.5 |
| Tween 20 | 10% (w/v) | 73±4 | 58±4 |
| PMSF | 10 mM | 1.1±0.2 | 0.5±0.9 |
| EDTA | 20 mM | 41±3 | 68±5 |
| iodoacetamide | 10 mM | 108±9 | 130±5 |
| DTT | 10 mM | 69±10 | 99±15 |
| urea | 6 M | 149±11 | 178±16 |
| Gdm·HCl[a] | 6 M | 26±1 | 88±4 |
| urea (70°C) | 6 M | 2.0±1.9 | 64±1 |
| Gdm·HCl (70°C) | 6 M | 0.7±1.2 | 9.1±0.3 |

[a] Gdm, guanidinium chloride

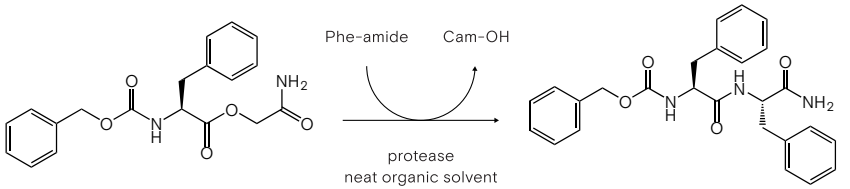
**TABLE 2.**

Effect of different agents on *Taq*Sbt and *Dg*Sbt activity. Enzymes were incubated for 1 h at 40°C or 70°C (where indicated) in 100 mM Hepes·NaOH buffer, pH 7.5, supplemented with the agent indicated. Remaining activities were measured in triplicate with *N*-suc-AAPF-*p*NA as the substrate.

Both subtilases tolerated the anionic surfactant SDS and non-ionic Tween 20 with some loss of activity, whereas a cationic surfactant inactivated the enzymes. Surprisingly, incubation with 6 M urea at 40°C activated both proteases, whereas the more thermostable *Taq*Sbt remained active after incubation with 6 M urea at 70°C, showing the extreme stability of the enzyme. This exceptional stability of the novel subtilases in the presence of high concentrations of EDTA, SDS, Tween 20, urea and guanidinium hydrochloride (Gdm•HCl) as compared to the engineered chaophilic protease[47] and hyperthermophilic protease Tk-SP[48] makes the enzymes interesting candidates for application under harsh process conditions.

## Application to peptide synthesis at high temperature in organic solvents

The possibility to use the proteases *Dg*Sbt and *Taq*Sbt for biocatalytic peptide synthesis under anhydrous conditions was investigated. A possible method for the preparation of dry enzyme powder is precipitation of the enzyme from buffer by an organic solvent[49].Initial synthetic reactions were carried out with enzyme precipitated with t-butanol as reported for subtilisin A32, and synthesis of Cbz-Phe-Phe-$NH_2$ was tested with Cbz-Phe-OCam ester as acyl donor and a H-Phe-$NH_2$ as nucleophile in 1:1.5 molar ratio in neat tetrahydrofuran (Figure 3). Prior to enzyme addition, the mixture was supplemented with dried molecular sieves (crushed, 3Å) to absorb any remaining water. Under these conditions, *Taq*Sbt gave 1% of synthetic product and *Dg*Sbt 11% of synthetic product after 24 h incubation. Thus, the dried enzymes were active in peptide synthesis under anhydrous conditions, which prompted us to test various alternative solvents.




Cbz-Phe-OCam

Cbz-Phe-Phe amide





**FIGURE 3.**

Protease-catalysed dipeptide synthesis (Cbz-Phe-Phe-$NH_2$) in neat organic solvent. Starting materials include neat organic solvent, 50 mM acyl donor in the form of the carboxamidomethyl ester, nucleophile amide (75 mM) 3Å molecular sieves and water-free protease preparation.

To examine which conditions are most suitable for dry enzyme preparation, several water-miscible organic solvents were used to prepare precipitated enzymes that were subsequently tested in synthetic reactions. In addition, enzyme co-lyophilized with sucrose was tested (Table 5.3). Hydrolytic activities of enzyme preparations, apart from enzyme precipitated with acetonitrile, were preserved when the dried enzymes were resuspended in buffer and re-tested, suggesting that little irreversible denaturation or inactivation occurs and overall activity is well maintained during preparation of anhydrous enzyme.

| Enzyme | Activity % after lyophilization[b] | | Activity % after precipitation | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | acetonitrile | | *t*-butanol | | isopropanol | | acetone | |
| | Syn | Hyd | Syn | Hyd | Syn | Hyd | Syn | Hyd | Syn | Hyd |
| *Dg*Sbt | 29 | 86 | 15 | 33 | 49 | 66 | 82 | 72 | 44 | 100 |
| *Taq*Sbt | 27 | 116 | 1 | 41 | 37 | 70 | 88 | 91 | 4 | 95 |

[a] Synthetic activities (Syn) were determined for Cbz-Phe-Phe-$NH_2$ synthesis in acetonitrile at 60°C as yields after 30 min. Retained hydrolytic activity (Hyd) was determined after redissolving the enzyme in buffer with the amidolytic assay. Hydrolytic activity of untreated enzyme was set to 100%.

[b] Lyophilization was performed after addition of 10 % sucrose (wt/wt of protein).

**TABLE 3.**

Synthetic performance of lyophilized enzyme and anhydrous enzyme preparations obtained by precipitation with organic solvents.[a]

Synthetic activity was observed with all enzyme preparations, with conversions after 30 min in the range of 1 to 88%, confirming that the novel thermostable proteases can act as efficient catalysts for peptide production in anhydrous media. Although acetonitrile was a good solvent for synthetic reactions (see below), it gave low activities when used for enzyme precipitation. Precipitation with acetone, which is often used for protein precipitation, also gave modest synthetic activity. The highest synthetic activity was observed with the isopropanol precipitated samples of the *Dg*Sbt and *Taq*Sbt proteases, with product synthesis up to 88% (Table 3). Drying of these isopropanol-precipitated and -rinsed enzyme preparations (IPREP) in a Speedvac gave a stable white powder (IPRED) that could be stored as active biocatalyst at room temperature. This enzyme was used in further peptide synthesis experiments. Water content of blank reactions (no enzyme) was 0.04%, and reactions with IPRED preparations added had 0.12–0.13% water, showing that lyophilized enzyme still contains some bound water.

High reaction temperatures can offer several advantages in enzymatic peptide synthesis, including better mass transfer, elevated substrate and product solubility, and increased reaction rates. Therefore, dipeptide synthesis in anhydrous acetonitrile was attempted at elevated temperature (60°C, 37°C for comparison), using IPREP as the catalyst. Again, highly efficient peptide synthesis was observed with *Taq*Sbt and *Dg*Sbt when monitored in time. Reactions reached 90–100% conversion to synthetic product, also at 60°C (Figure 4), which in case of *Taq*Sbt and *Dg*Sbt gave higher conversion rates than at 37°C. Since *Taq*Sbt has optimal activity at about 90°C, peptide synthesis was attempted also at 80°C; again full conversion to dipeptide product was observed. At 37°C, the initial rates of synthesis with the fastest thermostable subtilase (*Taq*Sbt) were lower than with subtilisin A, but the activity of *Taq*Sbt was enhanced 2.6-fold by increasing the temperature to 60°C. Although the activity was still lower than that of subtilisin A at 37°C, subtilisin A gave incomplete conversion which we attribute to thermal inactivation during the reaction (Figure 4).

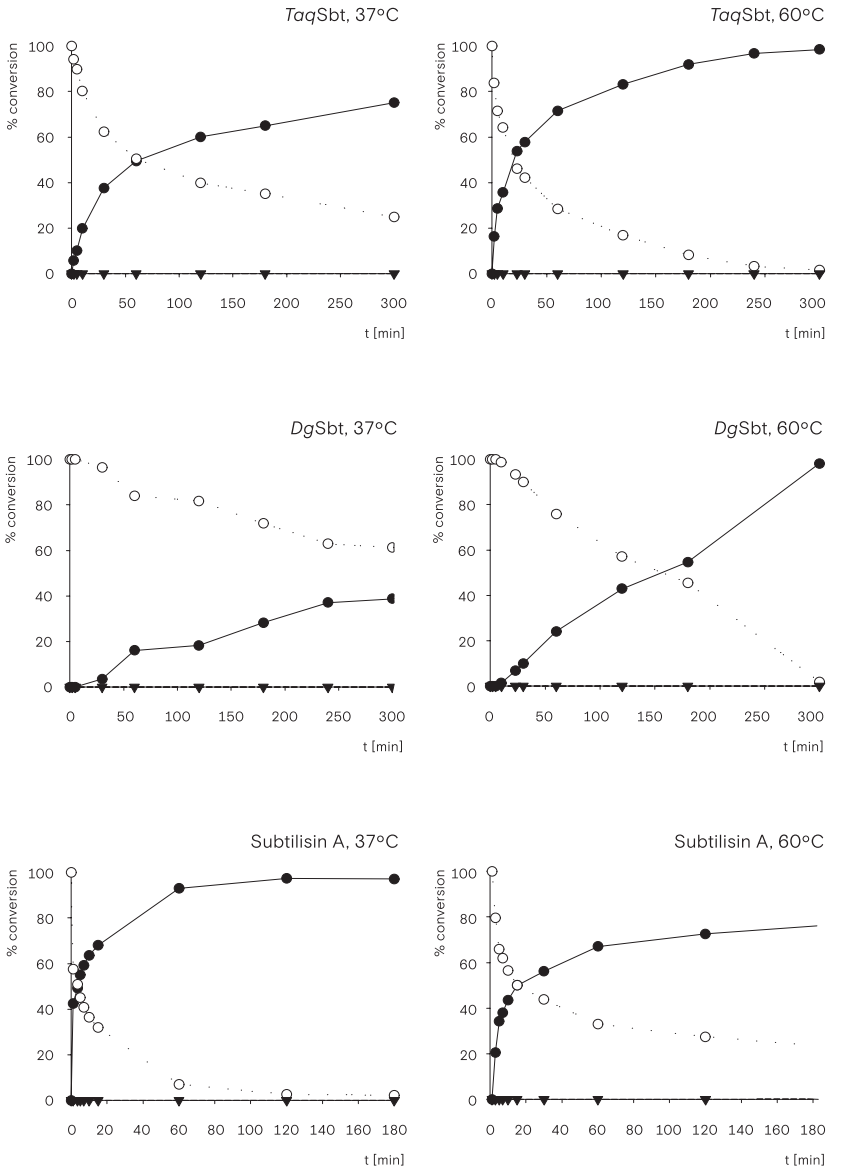



**FIGURE 4.**

Temperature effect on Cbz-Phe-Phe-NH2 synthesis using the thermostable *Taq*Sbt and *Dg*Sbt enzymes and subtilisin A. Reactions were performed in acetonitrile at 37°C and at 60°C as indicated in Materials and Methods. Symbols: ( ○ ), starting material Cbz-Phe-OCam; ( ● ), synthetic product Cbz-Phe-Phe amide; ( ▼ ), hydrolysis product Cbz-Phe-OH.

The solubility of a peptide in a solvent or buffer is mainly influenced by its polarity, which is determined by the amino acid composition, and its conformation. Highly hydrophobic solvents will decrease the solubility of peptides bearing polar groups. Peptide solubility can be enhanced by selecting appropriate solvent mixtures. For example, addition of the strongly solvating compound dimethylformamide (DMF) is known to increase solubility of peptides and DMF is used as a solvent in chemical solution- or solid-phase peptide synthesis. On the other hand, enzymes usually operate only in a restricted range of solvent hydrophobicities[50,51]. Therefore, the levels of hydrophilic solvents (logP < 2) in reaction mixtures are usually kept low. Moreover, in a recent study it was shown that stability of subtilisin A is inversely proportional to the amount of hydrophilic solvent in acetone/dodecane mixtures used in the reaction[52].

To investigate the effect of different cosolvents on peptide synthesis with the isolated thermostable proteases, we measured peptide coupling in solvent mixtures containing 40% DMF and 60% (v/v) of different cosolvents with varying logP values (Table 4). Reactions were done at high temperature (60 and 80°C). In these experiments, the product yields measured after 1h and 24 h were highest in case of acetonitrile-DMF mixtures, and lowest in tetrahydrofuran -DMF mixtures. The synthetic yield was not correlated to the logP value of the solvent added to DMF. With both enzymes, the best conversions were obtained with DMF mixtures containing acetonitrile (polar aprotic) or a polar protic solvent such t-butanol. No difference in performance or cosolvent preference was observed when using either fresh IPREP or dry IPRED powder (not shown).

| 40%DMF-60% solvent | b.p. (°C) | type | logP[b] | $logP_{mixture}$ | Synthetic product Cbz-Phe Phe-$NH_2$ formed (%)[a] | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | *Taq*Sbt (80°C) | | *Dg*Sbt (60°C) | |
| | | | | | 1h | 24h | 1h | 24h |
| Acetonitrile | 82 | polar aprotic | -0.39 | -0.42 | 32 | 53 | 37 | 90 |
| Dioxane | 101 | non polar | -0.25 | -0.35 | 3 | 11 | 21 | 76 |
| Isopropanol | 73 | polar protic | 0.17 | -0.18 | 1 | 11 | 4 | 92 |
| Tetrahydrofuran | 66 | polar aprotic | 0.47 | -0.04 | 2 | 3 | 7 | 45 |
| *t*-Butanol | 85 | polar protic | 0.58 | 0.02 | 12 | 47 | 26 | 91 |
| Toluene | 111 | non polar | 2.72 | 1.04 | 7 | 31 | 19 | 79 |
| Hexane | 69 | non polar | 3.76 | 1.54 | 8 | 14 | 11 | 54 |

**TABLE 4.**

Solvent screen for Cbz-Phe-Phe-amide synthesis. Reactions were done in the presence of 40% DMF at 60 and 80°C in mixtures containing 10 wt% of IPREP, corresponding to final concentrations of 11µM for *Taq*Sbt and 25 µM for *Dg*Sbt.

[a] The amount of hydrolytic product after 1 h reaction time was less than 3%, except in a reaction with TaqSbt in acetonitrile (17%). After 24 h, hydrolytic product was <10%, except in a reaction with *Dg*Sbt in tetrahydrofuran (13%) and *Taq*Sbt in acetonitrile (32%).

[b] Calculated using Advanced Chemistry Development Software V11.02 (ACD/Labs). For binary mixtures $logP_{mixture}$ values were calculated as reported by Laane et al.[50].

| Enzyme | DMF present (vol%) | | | | | |
|---|---|---|---|---|---|---|
| | 0 | 20 | 40 | 60 | 80 | 100 |
| | Syn (%) | Syn (%) | Syn (%) | Syn (%) | Syn (%) | Syn (%) |
| *Taq*Sbt | 50 | 37 | 21 | 8 | 2 | 0 |
| *Dg*Sbt | 52 | 25 | 24 | 12 | 6 | 1 |

**TABLE 5.**

DMF tolerance of *TaqSbt* and *DgSbt* during synthesis of Cbz-Phe-Phe-$NH_2$. Reactions were performed using 10 wt% of IPRED for 1 h at 60°C in an anhydrous acetonitrile–DMF cosolvent system. In 0.1 mL reaction the concentrations of active enzyme were 11 µM (TaqSbt) and 25 µM (DgSbt). Enzymatic hydrolysis was less than 10%.

In subsequent experiments, DMF-acetonitrile cosolvent mixtures were used to optimize the amount of DMF, and several levels of DMF were tested (Table 5). Although DMF reduces the enzyme activity, a range of DMF concentrations at elevated temperature were reasonably well tolerated by the thermophilic subtilases and only at concentrations exceeding 80% DMF (v/v) enzymes lost all activity.

## Substrate scope

In kinetically controlled peptide coupling reactions, the acyl donor is usually applied in an activated form, e.g. as an ester. A simple ester, such as a methyl ester, can be used with *Taq*Sbt and *Dg*Sbt[38], but the use of highly activated carbamoylmethyl (Cam, glycolamide) or trifluoroethyl (Tfe) esters can increase the rate of peptide synthesis[53] and expand the protease substrate scope[14]. To test the activity with different peptides, several reactions were attempted on small scale (Table 6). The acyl donors tested included peptide Cam esters of variable size, a single amino acid, a dipeptide and a pentapeptide. As mentioned above, excellent conversions to the dipeptide product Cbz-Phe-Phe-$NH_2$ was obtained with Cbz-Phe-OCam as acyl donor and Phe-$NH_2$ as the nucleophile (Figure 3). Hydrolysis was negligible. This well-accepted acyl donor with a hydrophobic residue of Phe at P1 position was below for reactions with sterically demanding nucleophiles. When more difficult acyl donors bearing a Pro or Ile at the P1 position were attempted, only *Taq*Sbt gave reasonable conversion, with formation of the synthetic products Cbz-Val-Pro-Val-$NH_2$ and Cbz-Val-Ile-Val-$NH_2$ up to 35% and 18%, respectively (Table 6). The acyl donor Ac-Asp-Phe-Ser-Lys-Leu-OCam is insoluble in pure acetonitrile. By dissolving it in solvent mixture with 50% (*v/v*) DMF, which is tolerated by both subtilases, a coupling reaction was possible (Table 6). *Dg*Sbt was more efficient in these coupling reactions with long acyl donors and gave hexameric product, heptameric product and octameric product in yields of 55 – 88%. *Taq*Sbt performed best in heptamer synthesis. The reactions with these long acyl donors were only possible in the presence of DMF as cosolvent, which solubilized the substrate. Inspired by the broad substrate scope with acyl donors, we tested nucleophiles that vary in size and composition, including a peptide with sterically demanding amino acids and a peptide containing a D-amino acid at P2' position. Nucleophilic peptides with hydrophobic residues at P1′(Phe, Ala) or P2′ (Phe), such as H-Phe-$NH_2$ or H-Ala-Phe-$NH_2$, were well accepted (Tables 4 and 6). They could be coupled to a short acyl donor like Cbz-Phe-OCam with 82% formation of synthetic product (Cbz-Phe-Ala-Phe-$NH_2$) using *Dg*Sbt. The longer nucleophiles were also coupled to a pentameric acyl donor (Ac-Asp-Phe-Ser-Lys-Leu-OCam), and the highest conversion to octamer (84%) was observed in case of *Dg*Sbt. Even a sterically hindered peptide with proline at P1' was accepted as a nucleophile if used at high concentration. High concentrations of nucleophile may not only outcompete remaining water (if present), but also increase the rate of formation of the coupling product if deacylation is slow. *Taq*Sbt catalyzed the reaction of Cbz-Phe-OCam as acyl donor with H Pro-$NH_2$ with good conversion to the dipeptide Cbz-Phe-Pro-$NH_2$ (Table 6). A nucleophile containing a D-amino acid at the P2′ position was accepted by both enzymes and could be coupled to Cbz-Phe-OCam resulting in tetrapeptide product, again in good yield.

| Acyl donor (mM) | Nucleophile (mM) | Synthetic product | *Dg*Sbt | | | *Taq*Sbt | | |
|---|---|---|---|---|---|---|---|---|
| | | | Conv (%) | Syn (%) | Hyd (%) | Conv (%) | Syn (%) | Hyd (%) |
| Cbz-Val-Ile-OCam (2.5) | H-Val-$NH_2$ (3.75) | Cbz-Val-Ile-Val-$NH_2$ [a] | <1 | <1 | <1 | 18 | 18 | <1 |
| Cbz-Val-Pro-OCam (25) | H-Val-$NH_2$ (3.75) | Cbz-Val-Pro-Val-$NH_2$ [b] | 2.5 | 2.5 | <1 | 35 | 35 | <1 |
| Cbz-Phe-OCam (50) | H-Ala-Phe-$NH_2$ (75) | Cbz-Phe-Ala-Phe-$NH_2$ [c] | 88 | 82 | 6 | 76 | 68 | 8 |
| Cbz-Phe-OCam (50) | H-Pro-$NH_2$ (250) | Cbz-Phe-Pro-$NH_2$ [d] | 44 | 38 | 6 | 83* | 73 | 6 |
| Cbz-Phe-OCam (5) | H-Ala-(D)Leu-Arg-$NH_2$ (7.5) | Cbz-Phe-Ala-(D) Leu-Arg-$NH_2$ [e] | 99 | 76 | 23 | 95* | 88 | 4 |
| Ac-Asp-Phe-Ser-Lys- Leu-OCam (20) | H-Ala-$NH_2$ (30) | Ac-Asp-Phe-Ser-Lys-Leu-Ala-$NH_2$ [f] | 88 | 73 | 15 | 26 | 26 | <1 |
| Ac-Asp-Phe-Ser-Lys- Leu-OCam (20) | H-Ala-Phe-$NH_2$ (30) | Ac-Asp-Phe-Ser-Lys-Leu-Ala-Phe- $NH_2$ [g] | 55 | 55 | <1 | 75 | 75 | <1 |
| Ac-Asp-Phe-Ser-Lys- Leu-OCam (20) | H-Ala-Phe-Ala-$NH_2$ (30) | Ac-Asp-Phe-Ser-Lys-Leu-Ala-Phe-Ala-$NH_2$ [h] | 89 | 84 | 5 | 16 | 16 | <1 |

Reaction conditions: [a] 40% DMF, 15 mg IPRED, 150 µL reaction, 100 mg/mL Å3 sieves, $Et_3N$ 1.5 µL, 72 h; | [b] 40% DMF, 15 mg IPRED, 150 µL reaction, 100 mg/mL Å3 sieves, $Et_3N$ 1.5 µL, 72 h; | [c] 40% DMF, 15 mg IPRED, 150 µL reaction, 100 mg/mL Å3 sieves, $Et_3N$ 5 µL, 24 h; | [d] 40% DMF, 10 mg IPRED, 100 µL reaction, 45 mg/mL Å3 sieves, $Et_3N$ 5 µL, 24 h; |[e] 40% DMF, 10 mg IPRED, 100 µL reaction, 50 mg/mL Å3 sieves, $Et_3N$ 5 µL, 96 h for *Taq*Sbt and 24 h for *Dg*Sbt; | [f] 20 mM acyl donor, 30 mM nucleophile, 50% DMF, 10 mg IPRED, 100 µL reaction, 20 mg/mL Å3 sieves, $Et_3N$ 5 µL, 24 h; | [g] 50% DMF, 110 mg IPRED, 500 µL reaction, 20 mg/mL Å3 sieves, $Et_3N$ 5 µL, 24 h; h 50% DMF, 10 mg IPRED, 100 µL reaction, 20 mg/mL Å3 sieves, $Et_3N$ 5 µL, 24 h.
* Chemical hydrolysis observed (up to 4%). 10 mg of *Taq*Sbt IPRED contained 0.2 mg of active subtilase and 10 mg of *Dg*Sbt IPRED contained 0.5 mg of active subtilase determined by active site titration

**TABLE 6.**

Substrate scope of *Taq*Sbt and *Dg*Sbt subtilase.Peptide coupling reactions were performed at 60°C in neat organic solvent with DMF as indicated below.

In most cases, hydrolysis was very low, like in the progress curves shown in Figure 5. Water content was determined and indeed appeared low. For example, in the heptamer (Ac-Asp-Phe-Ser-Lys-Leu-Ala-Phe-$NH_2$) synthesis reactions, the blank reaction (no enzyme) contained 0.04% water and an incubation mixture with enzyme added had 0.12-0.13% water, regardless of the enzyme identity. No difference in water content was observed between reactions with fresh *Dg*Sbt IPRED or an IPRED preparation that had been stored for 1 month. The ternary solvent system is too complex for calculating water activity using the equations reported by Bell *et al.*[54], but aw may be in the range of 0.01-0.03. However, even when carefully applying dry conditions, including water removal by molecular sieves, some hydrolysis of the Cam ester acyl donor was still observed in case of the more challenging substrates, especially with a D-amino acid in the nucleophile (Table 6). This could be due to water strongly bound to the enzyme and not absorbed by the molecular sieves, which may be released and react during the incubation.

## Two enzyme *N→C* peptide elongation strategy: proof of principle

In another set of experiments, we further explored the applicability of novel thermostable subtilases in *N→C* peptide elongation reaction in combination with an amidase from *Stenotrophomonas maltophilia* (PAM). In neat organic solvent, the amidase can perform *C*-terminal deprotection-reactivation by catalyzing the conversion of a peptide amide to the methyl ester. The latter can be coupled to as nucleophilic peptide or amino acid in a kinetically controlled reaction step catalyzed by a protease (Figure 5).





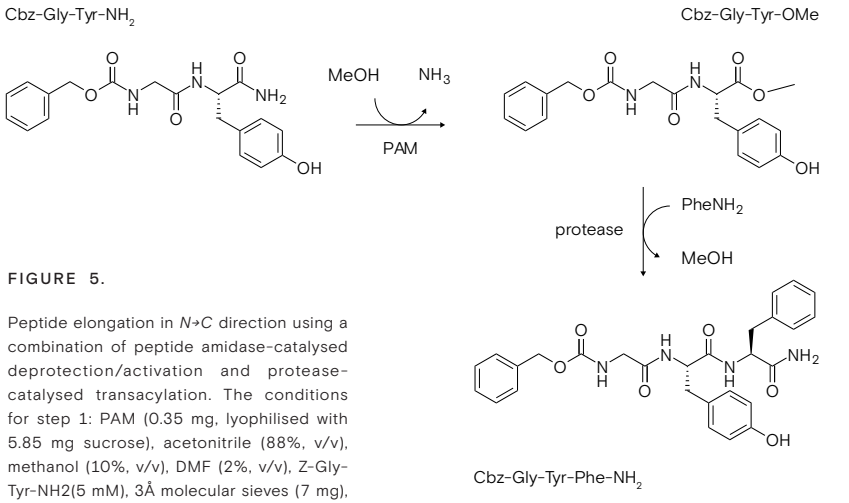



**FIGURE 5.**

Peptide elongation in *N→C* direction using a combination of peptide amidase-catalysed deprotection/activation and protease-catalysed transacylation. The conditions for step 1: PAM (0.35 mg, lyophilised with 5.85 mg sucrose), acetonitrile (88%, v/v), methanol (10%, v/v), DMF (2%, v/v), Z-Gly-Tyr-NH2(5 mM), 3Å molecular sieves (7 mg), 5Å molecular sieves (3.5 mg), total volume 0.25 mL, 30°C, 16. Conditions for protease catalysed coupling are described under Material and Methods.

For this two-enzyme reaction, PAM was used first for the conversion of the acyl donor Cbz-Gly-Tyr-$NH_2$ to the methyl ester. Next, the two new thermostable proteases *Dg*Sbt and *Taq*Sbt described here were used for coupling the obtained Z-Gly-Tyr-OMe with Phe-$NH_2$ as the nucleophile (Figure 5). The peptide amidase was removed by centrifugation after the first step, after which the protease *Dg*Sbt was added directly to the reaction mixture. Although the reaction rate was low, the conversion reached a notable value of 76%. In another experiment, enzymatically obtained and purified Z-Gly-Tyr-OMe was used as a substrate in a subsequent *Taq*Sbt protease-catalysed coupling reaction yielding 77% of Z-Gly-Tyr-Phe-$NH_2$ product over 6 days. After removal of the protease from the reaction mixture, the product could in principle be used in another elongation step.

## Conclusion

Two novel thermostable subtilases (*Dg*Sbt and *Taq*Sbt) were biochemically characterized and their applicability in chemo-enzymatic peptide segment condensation reactions in neat organic solvents, especially at high temperatures and with DMF as cosolvent, was demonstrated. Such conditions can improve substrate solubility, minimize the need for side chain group protection, and increase reaction rates. A simple and convenient enzyme preparation method was developed (IPREP/IPRED). It combines three steps: enzyme precipitation by alcohol (isopropanol); water removal by solvent washing; and precipitation in the presence of a high salt concentration that may increase activity of the enzyme preparation in anhydrous organic solvent. Drying of IPREP enzyme to a powder (IPRED) that can be stored at room temperature makes it convenient for use on laboratory scale. We observed only 30% reduction in activity after 1 month storage at room temperature. As shown, such IPRED preparations were efficient in 1+1 and 1+2 peptide coupling. At increased DMF levels, we could use higher substrate concentrations and also perform coupling reactions with substrates that are poorly soluble in other commonly used solvents.

*Dg*Sbt and *Taq*Sbt subtilases expand the scope of enzymatic peptide coupling reactions in neat organic solvents to high-temperature and high-DMF reaction conditions. Proline is accepted at the P1 and P1' position, and also a D-amino acid is accepted in the S2' pocket. We also found that the use of thermostable subtilases at high-temperature and high-DMF concentration makes it possible to synthesize an octapeptide. Higher conversions can possibly be achieved by reaction optimization by means of more precise control of water activity, by addition of larger amounts of enzyme, or by applying prolonged incubation times.

In addition, the *Dg*Sbt and *Taq*Sbt subtilases was applied in combination with an amidase in a peptide elongation reaction in the *N→C* direction on preparative scale in neat organic solvent.

**Acknowledgments**

This work was supported by the program Integration of Biosynthesis and Organic Synthesis (IBOS-2; project 053.63.014), funded by The Netherlands Organization for Scientific Research (NWO). The authors would like to thank dr. Fabrizia Fusetti and dr. Alicja Filipowicz for their assistance with mass spectrometry analysis.

**Author contribution**

Experimental work on proteases was performed by AT. TN synthesized and provided peptide substrates. IA did all experimental work with the amidase. BW provided analytical support. AT and DBJ wrote the paper. PJLMQ and DBJ supervised the work.

**Supporting information**

The supplementary material is available at

http://www.sciencedirect.com/science/article/pii/S0141022915000472.

# References


01 — Thayer, A. Improving peptides. *Chem. Eng. News* **89**, 13–20 (2011).

02 — Thayer, A. M. Making peptides at large scale. *Chem. Eng. News* **89**, 9–12 (2011).

03 — Lax, R. The future of peptide development in the pharmaceutical industry. *PharManufacturing Int. Pept. Rev.* 10–15 (2010).

04 — Mine, Y., Li-Chan, E. & Jiang, B. *Bioactive proteins and peptides as functional foods and nutraceuticals.* (Wiley-Blackwell, 2010).

05 — Bordusa, F. Proteases in organic synthesis. *Chem. Rev.* **102**, 4817–68 (2002).

06 — Bordusa, F. Substrate mimetics in protease catalysis: characteristics, kinetics, and synthetic utility. *Curr. Protein Pept. Sci.* **3**, 159–180 (2002).

07 — Youshko, M. I., Chilov, G. G., Shcherbakova, T. A. & Svedas, V. K. Quantitative characterization of the nucleophile reactivity in penicillin acylase-catalyzed acyl transfer reactions. *Biochim. Biophys. Acta* **1599**, 134–40 (2002).

08 — Youshko, M. I., Van Langen, L. M., de Vroom, E., van Rantwijk, F., Sheldon, R. A. & Švedas, V. K. . Highly efficient synthesis of ampicillin in an 'aqueous solution-precipitate' system: Repetitive addition of substrates in a semicontinuous process. *Biotechnol. Bioeng.* **73**, 426–430 (2001).

09 — Kitaguchi, H. in *Enzymatic reactions in organic media* (eds. Koskinen, A. & Klibanov, A. M.) 224–243 (Chapman&Hall, 1996).

10 — Clapés, P., Torres, J. L. & Adlercreutz, P. Enzymatic peptide synthesis in low water content systems: preparative enzymatic synthesis of [Leu]- and [Met]-enkephalin derivatives. *Bioorg. Med. Chem.* **3**, 245–55 (1995).

11 — Klibanov, A. M. Improving enzymes by using them in organic solvents. *Nature* **409**, 241–246 (2001).

12 — Nuijens, T., Cusan, C., van Dooren, T. J. G. M., Moody, H. M., Merkx, R., Kruijtzer, J. A. W., Rijkers, D. T. S., Liskamp, R. M. J. & Quaedflieg, P. J. L. M. Fully enzymatic peptide synthesis using C-terminal tert-butyl ester interconversion. *Adv. Synth. Catal.* **352**, 2399–2404 (2010).

13 — Nuijens, T., Piva, E., Kruijtzer, J. A. W., Rijkers, D. T. S., Liskamp, R. M. J. & Quaedflieg, P. J. L. M. Fully enzymatic N→C-directed peptide synthesis using C-terminal peptide α-carboxamide to ester interconversion. *Adv. Synth. Catal.* **353**, 1039–1044 (2011).

14 — Nuijens, T., Schepers, A. H. M., Cusan, C., Kruijtzer, J. A. W., Rijkers, D. T. S., Liskamp, R. M. J. & Quaedflieg, P. J. L. M. Enzymatic fragment condensation of side chain-protected peptides using Subtilisin A in anhydrous organic solvents: A general strategy for industrial peptide synthesis. *Adv. Synth. Catal.* **355**, 287–293 (2013).

15 — Koskinen, A. & Klibanov, A. M. *Enzymatic reactions in organic media*. (blackie academic &Professional, an imprint of Chapmann &Hall, 1996).

16 — Vossenberg, P., Beeftink, H. H., Nuijens, T., Quaedflieg, P. J. L. M., Cohen Stuart, M. A. & Tramper, J. Performance of Alcalase formulations in near dry organic media: Effect of enzyme hydration on dipeptide synthesis. *J. Mol. Catal. B Enzym.* **78**, 24–31 (2012).

17 — Miyazawa, T., Hiramatsu, M., Murashima, T. & Yamada, T. *Bacillus licheniformis* protease-catalyzed peptide synthesis via the kinetically controlled approach using the carbamoylmethyl ester as an acyl donor in anhydrous acetonitrile. *Lett. Pept. Sci.* **9**, 173–177 (2002).

18 — Richards, A. O., Gill, I. S. & Vulfson, E. N. Continuous enzymatic production of oligopeptides: Synthesis of an enkephalin pentapeptide in a multistage bioreactor. *Enzyme Microb. Techno*l. **15**, 928–935 (1993).

19 — Schulze, B. & Klibanov, A. M. Inactivation and stabilization of subtilisins in neat organic solvents. *Biotechnol. Bioeng.* **38**, 1001–1006 (1991).

20 — Nuijens, T., Cusan, C., Schepers, A. H. M., Kruijtzer, J. A. W., Rijkers, D. T. S., Liskamp, R. M. J. & Quaedflieg, P. J. L. M. Enzymatic synthesis of activated esters and their subsequent use in enzyme-based peptide synthesis. *J. Mol. Catal. B Enzym.* **71**, 79–84 (2011).

21 — Nyfeler, R. in *Methods in Molecular Biology,* Vol 35, *Peptide Synthesis Protocols* (eds. Pennington, M. . & Dunn, B. M.) **35**, 303–316 (Humana Press Inc., 1994).

22 — Gololobov M.Yu, Stepanov, V. M., Voyushina, T. L., Morozova, I. P. & Adlercreutz, P. Side reactions in enzymatic peptide synthesis in organic media: effects of enzyme, solvent, and substrate concentrations. *Enzyme Microb. Technol.* **16**, 522–8 (1994).

23 — Margolin, A. L., Tai, D. & Klibanov, A. M. Incorporation of D-amino acids into peptides via enzymatic condensation in organic solvents. *J. Am. Chem. Soc.* **109**, 7885–7887 (1987).

24 — Chen, K. & Arnold, F. H. Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 5618–22 (1993).

25 — Chen, K. Q. & Arnold, F. H. Enzyme engineering for nonaqueous solvents: random mutagenesis to enhance activity of subtilisin E in polar organic media. *Biotechnol. Nat. Publ. Co.* **9**, 1073–1077 (1991).

26 — Zhong, Z. Liu, J. L., Dinterman, L. M., Finkelman, M. A. J., Mueller, W. T., Rollence, M. L., Whitlow, M. & Wong, C.-H. Engineering subtilisin for reaction in dimethylformamide. *J. Am. Chem. Soc.* **100**, 683–684 (1991).

27 — Liszka, M. J., Clark, M. E., Schneider, E. & Clark, D. S. Nature versus nurture: developing enzymes that function under extreme conditions. *Annu. Rev. Chem. Biomol. Eng.* **3**, 77–102 (2012).

28 — Floor, R. J., Wijma, H. J., Colpa, D. I., Ramos-Silva, A., Jekel, P. A., Szymański, W., Feringa, B. L., Marrink, S. J. & Janssen, D. B. Computational library design for increasing haloalkane dehalogenase stability. *Chembiochem* 1–14 (2014).




CHAPTER — 05 — REFERENCES

29 — de Beer, R. J. A. C., Nuijens, T., Wiermans, L., Quaedflieg, P. J. L. M. & Rutjes, F. P. J. T. Improving the carboxyamidomethyl ester for subtilisin A-catalysed peptide synthesis. *Org. Biomol. Chem.* **10**, 6767–75 (2012).

30 — Terada, I., Kwon, S. T., Miyata, Y., Matsuzawa, H. & Ohta, T. Unique precursor structure of an extracellular protease, aqualysin I, with $NH_2$- and COOH-terminal pro-sequences and its processing in *Escherichia coli*. *J. Biol. Chem.* **265**, 6576–81 (1990).

31 — Partridge, J., Halling, P. J. & Moore, B. D. Practical route to high activity enzyme preparations for synthesis in organic media subtilisin Carlsberg and α-chymotrypsin gives 1000-fold. *Chem. com* 841–842 (1998).

32 — Chen, S., Chen, S. & Wang, K. Kinetically controlled peptide bond formation in anhydrous alcohol catalyzed by the industrial protease Alcalase. **188**, 55–58 (1992).

33 — Krieger, E., Koraimann, G. & Vriend, G. Increasing the precision of comparative models with YASARA NOVA a self-parameterizing force field. *Proteins Struct. Funct. Genet.* **47**, 393–402 (2002).

34 — Toplak, A., Wu, B., Fusetti, F., Quaedflieg, P. J. L. M. & Janssen, D. B. Proteolysin, a novel highly thermostable and cosolvent-compatible protease from the thermophilic bacterium *Coprothermobacter proteolyticus*. *Appl. Environ. Microbiol.* **79**, 5625–32 (2013).

35 — Ericsson, U. B., Hallberg, B. M., Detitta, G. T., Dekker, N. & Nordlund, P. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal. Biochem.* **357**, 289–98 (2006).

36 — Apweiler, R. The universal protein resource (UniProt) in 2010. *Nucleic Acids Res.* **37**, D169–D174 (2009).

37 — Fenyo, D., Wang, Q., DeGrasse, J. A., Padovan, J. C., Cadene, M. & Chait, B. MALDI sample preparation: the ultra thin layer method. *Journal of visualized experiments : JoVE* 192 (2007).

38 — Arif, M. I., Toplak, A., Szymanski, W., Feringa, B. L., Nuijens, T., Quaedflieg, P. J. L. M., Wu, B. & Janssen, D. B. One-step C -terminal deprotection and activation of peptides with peptide amidase from *Stenotrophomonas maltophilia* in neat organic solvent. *Adv. Synth. Catal.* **356**, 2197–2202 (2014).

39 — Matsuzawa, H., Tokugawa, K., Hamaoki, M., Mizoguchi, M., Taguchi, H., Terada, I., Kwon, S. T. & Ohta, T. Purification and characterization of aqualysin I (a thermophilic alkaline serine protease) produced by *Thermus aquaticus* YT-1. *Eur. J. Biochem.* **171**, 441–7 (1988).

40 — Peek, K., Daniel, R. M., Monk, C., Parker, L. & Coolbear, T. Purification and characterization of a thermostable proteinase isolated from *Thermus* sp. strain Rt41A. *Eur. J. Biochem.* **207**, 1035–44 (1992).

41 — Kim, D. W., Lee, Y. C. & Matsuzawa, H. Role of the COOH-terminal pro-sequence of aqualysin I (a heat-stable serine protease) in its extracellular secretion by *Thermus thermophilus*. *FEMS Microbiol. Lett.* **157**, 39–45 (1997).







42 — Kurosaka, K., Ohta, T. & Matsuzawa, H. A 38 kDa precursor protein of aqualysin I (a thermophilic subtilisin-type protease) with a C-terminal extended sequence: Its purification and in vitro processing. *Mol. Microbiol.* **20**, 385–389 (1996).

43 — Siezen, R. J. & Leunissen, J. a. Subtilases: the superfamily of subtilisin-like serine proteases. *Protein Sci.* **6**, 501–23 (1997).

44 — Sakaguchi, M., Takezawa, M., Nozawa, K., Kusakawa, T., Nagasawa, T., Sugahara, Y. & Kawakita, M. Role of disulphide bonds in a thermophilic serine protease aqualysin I from *Thermus aquaticus* YT-1. *J. Biochem.* **143**, 625–632 (2008).

45 — Lin, S. J., Yoshimura, E., Sakai, H., Wakagi, T. & Matsuzawa, H. Weakly bound calcium ions involved in the thermostability of aqualysin I, a heat-stable subtilisin-type protease of *Thermus aquaticus* YT-1. *Biochim. Biophys. Acta* **1433**, 132–8 (1999).

46 — Tanaka, T., Matsuzawa, H. & Ohta, T. Substrate specificity of Aqualysin I, a bacterial thermophilic alkaline serine protease from *Thermus aquaticus* YT-1; Comparison with proteinase K, subtilisin BPN' and subtilisin Carlsberg. *Biosci. Biotechnol. Biochem.* **62**, 2161–2165 (1998).

47 — Li, Z., Roccatano, D., Lorenz, M. & Schwaneberg, U. Directed evolution of subtilisin E into a highly active and guanidinium chloride- and sodium dodecylsulfate-tolerant protease. *ChemBioChem* **13**, 691–9 (2012).

48 — Foophow, T., Tanaka, S., Koga, Y., Takano, K. & Kanaya, S. Subtilisin-like serine protease from hyperthermophilic archaeon *Thermococcus kodakaraensis* with *N*- and *C*-terminal propeptides. *Protein Eng. Des. Sel.* **23**, 347–55 (2010).

49 — Roy, I. & Gupta, M. N. Preparation of highly active α-chymotrypsin for catalysis in organic media. *Bioorg. Med. Chem.* Lett. **14**, 2191–3 (2004).

50 — Laane, C., Boeren, S., Vos, K. & Veeger, C. Rules for optimization of biocatalysis in organic solvents. *Biotechnol. Bioeng.* **102**, 2–8 (2009).

51 — Halling, P. J. Thermodynamic predictions for biocatalysis in nonconventional media: theory, tests, and recommendations for experimental design and analysis. *Enzyme Microb. Technol.* **16**, 178–206 (1994).

52 — Choi, Y. S. & Yoo, Y. J. A hydrophilic and hydrophobic organic solvent mixture enhances enzyme stability in organic media. *Biotechnol. Lett.* **34**, 1131–5 (2012).

53 — Miyazawa, T., Ensatsu, E., Yabuuchi, N., Yanagihara, R. & Yamada, T. Superiority of the carbamoylmethyl ester as an acyl donor for the kinetically controlled amide-bond formation mediated by α-chymotrypsin. *J. Chem. Soc. Perkin Trans.* **1** 390–395 (2002).

54 — Bell, G., Janssen, A. E. & Halling, P. J. Water activity fails to predict critical hydration level for enzyme activity in polar organic solvents: Interconversion of water concentrations and activities. *Enzyme Microb. Technol.* **20**, 471–477 (1997).

CHAPTER — 06

# Characterization of a novel organic cosolvent-tolerant protease from *Pseudomonas mendocina* ymp

Ana Toplak[a,b], Marcelo F. Masman[a], Timo Nuijens[b], Marleen Otzen[a], Peter J. L. M. Quaedflieg[c] and Dick B. Janssen[a]

[a] Biochemical Laboratory, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, 9747 AG Groningen, the Netherlands
[b] EnzyPep B.V., Urmonderbaan 22,6167 RD Geleen, the Netherlands
[c] DSM Innovative Synthesis, 6167 RD Geleen, the Netherlands

Part of this chapter has been published in:
Toplak A, Nuijens T, Quaedflieg PJML, Wu B, Janssen DB. Peptide synthesis in neat organic solvents with novel thermostable protease. Enz Microb Technol (2015) 73-74: 20-28.

Arif MI, Toplak A, Szymanski W, Feringa BL, Nuijens T, Quaedflieg PJML, Wu B, Janssen DB. One-step C -terminal deprotection and activation of peptides with peptide amidase from *Stenotrophomonas maltophilia* in neat organic solvent. Adv Synth Catal (2014) 10: 2197-2202

## Abstract

Enzymes with tolerance towards organic solvents could expand the applicability of biocatalysts in various industrial processes. Here, we report recombinant production in *E. coli,* purification and biochemical characterization of a novel serine protease from *Pseudomonas mendocina*. This protease, termed *Pm*Sbt, exhibited remarkable stability in aqueous solutions containing DMSO and DMF up to 50% *v/v* and was applied in the chemoenzymatic synthesis of a hexameric segment of the PHM27 bioactive peptide.






## Introduction

Peptides receive increasing attention as pharmaceuticals for the treatment of infectious diseases, metabolic disorders and for use in cancer therapy[1]. Other applications are in the food and cosmetics industries[2–5]. The development of such applications requires efficient synthetic methods, for which chemoenzymatic peptide synthesis is frequently considered. Here, a protease is used in the synthetic direction for coupling of peptide segments that are obtained by solution- or solid-phase peptide synthesis (SPPS). With enzymes as catalysts of peptide bond formation, the use of organic solvents may offer advantages such as increased substrate solubility, shift of thermodynamic equilibrium to the synthetic product, suppression of water-dependent side reactions, alternation of substrate specificity and enantioselectivity, and elimination of microbial contamination[6].

Accordingly, engineering of a DMF-tolerant protease has been attempted[7,8] for application under conditions that improve substrate solubility and suppress undesired hydrolytic reactions. In addition, proteases that possess tolerance to organic solvents are of interest for application in peptide synthesis. Such enzymes can be found in extremophiles, i.e. thermophiles and halophiles, and in organic-solvent tolerant microorganisms[9]. Soil, waste water and marine habitats are rich sources of organic-solvent tolerant microorganisms[10]. Many of these microorganisms belong to the *Pseudomonas* and *Bacillus genera* and are able to grow in the presence of organic solvents[6,11]. Several organic-solvent tolerant metalloproteases were isolated from these organisms and were used in peptide synthesis under thermodynamic control, including pseudolysin,[12] originating from *Pseudomonas aeruginosa* PST-01, as well as recently described metalloproteases from *Pseudomonas aeruginosa* PT121 and *Bacillus cereus* WQ9-2[13,14]. Unlike metalloproteases, the serine or cysteine proteases can synthesize peptides under kinetic

control, which enables higher conversion and requires less enzyme[15]. Organic-solvent compatible serine or cysteine proteases could expand the applicability of chemoenzymatic peptide synthesis.

During genome mining of DNA sequence databases for genes encoding novel subtilisin-like peptidases, we identified a putative serine protease gene in *Pseudomonas mendocina* strain ymp[16]. The recombinant enzyme *Pm*Sbt showed remarkable stability in aqueous solutions containing DMSO and DMF. In this Chapter we describe the biochemical characterization of *Pm*Sbt, a novel serine protease originating from *P. mendocina* strain ymp that is stable in organic cosolvents. In addition, we performed bioinformatic studies to elucidate *Pm*Sbt properties, and we tested the enzyme in peptide synthesis.

## Materials and methods

**Chemicals and reagents —** Bacterial strain and plasmids are listed in Chapter 3. Peptide amides were obtained from Sigma-Aldrich and Bachem (Bubendorf, CH). C-terminal carbamoylmethyl (Cam, glycolamide) esters of peptides were synthesized by SPPS as reported by de Beer *et al.*[17]. Sypro Orange dye was from Molecular Probes (Life Technologies, Carlsbad, CA, USA).

**Cloning of the genes —** The gene coding for the YP_001189588 protein was amplified from genomic DNA of *Pseudomonas mendocina* ymp with a modified start codon to incorporate an NdeI restriction site and with the stop codon omitted and an HindIII restriction site introduced downstream the structural gene. Cloning is described in Chapter 3. The gene coding for the truncated protein *Pm*Sbtdel (PMdel) was constructed using PCR with a suitable reverse primer with Arg420 as the last amino acid before the linker and a hexahistidine tag and cloned as mentioned before. The variants *Pm*Sbt CxA (x=154, 247, 370, 414, 479, 490) with the respective cysteine replaced by an alanine were constructed using QuikChange (Stratagene). The sequences of recombinant plasmids were confirmed by sequencing of both strands (GATC, Konstanz, Germany).





**Enzyme production —** Plasmids pBADMycHisA(NdeI)-PM (AmpR) and pBADMycHisA(NdeI)-PMdel (AmpR) constructed for the recombinant protease production were transformed to *E. coli* C43(DE3). The transformants grown on Luria Bertani (LB) agar plates containing 50 µg/mL ampicillin at 37°C for 16 h were picked and inoculated into 5 mL of LB broth containing 0.1% glucose and antibiotic. After 16 h of incubation at 37°C, 0.1 % (v/v) of the precultures were inoculated to 1 liter of Terrific broth (12 g/l tryptone, 24 g/l yeast extract, 4 mL glycerol supplemented with 10% solution of 0.17 M $KH_2PO_4$ and 0.72 M $K_2HPO_4$) containing 50 µg/mL ampicillin and cells were grown until late log-phase at 37°C. Protein synthesis was induced at 17°C by addition of 0.25% (w/v) L-(+) arabinose. The cells were grown for another 72 h at 17°C with constant shaking at 200 rpm. Proteolytic activity was detected in the growth medium after 72 h.

For isolating cellular enzymes, the cells were harvested (Beckman, JLA 10.5 rotor, 4°C, 15 min, 6,000 x g), washed and suspended in buffer A (25 mM Hepes/NaOH pH 7.5, 1 mM $CaCl_2$, 20% glycerol) and disrupted by sonication. After centrifugation (Beckman, JA-17rotor, 4°C, 45 min, 30, 000 x g) the clear cell-free extract was loaded onto HisTrap (GE Healthcare). The protein was eluted using 125 mM imidazole. Active protein fractions were collected and concentrated with an Amicon YM10 ultrafiltration membrane (Millipore, Billerica, MA, USA) followed by buffer exchange to buffer A.

Protease in the culture medium was precipitated using ammonium sulfate (70% sat. at 4°C) and incubated for 1 h at 4°C, followed by a centrifugation step (8,000 rpm, 30 min, JLA 9.1000, 4°C). The protein pellet was resuspended in buffer A and desalted on HiPrep column (GE Healthcare). The desalted fraction was loaded onto HisTrap column and protein was purified as described above. Alternatively, the culture medium carrying proteolytic activity was concentrated with an Amicon YM10 ultrafiltration membrane (Millipore, Billerica, MA, USA), followed by subsequent buffer exchange to buffer A and purification on the HisTrap column (GE Healthcare). Protein concentrations were determined by measuring absorption at 280 nm using bovine serum albumin (BSA) as standard, and purity was determined by SDS-PAGE analysis. For further analysis, active site concentrations were determined by PMSF titration as reported previously[18]. The processed pure protein was stored at -20 °C.

**Enzyme assays —** General proteolytic activity of *E. coli* cells carrying gene for *Pm*Sbt and its variants was tested on agar plates containing 1% skim milk and appropriate antibiotic. The proteolytic activity was observed as

a clear halo surrounding the colonies. Amidolytic activity was quantified by following the release of *p*-nitroaniline from the standard proteolytic substrate *N*-succinyl-Ala-Ala-Pro-Phe *p*-nitroanilide (*N*-suc-AAPF-pNA) at 40°C in 100 mM Hepes·NaOH, pH 7.5, containing 1 mM $CaCl_2$, 10 % (*v/v*) DMSO and substrate (0.1-5 mM)[19]. The pH optimum was determined in triplicate using phosphate buffer (0.1 M, pH range 5.7-8.0), Tris buffer (0.1M, pH 6.8-9.4) and carbonate buffer (0.1 M, pH 9.0-10.8) in amidolytic assays. The pH was adjusted at the assay temperature (40°C). In addition, azocasein assays were used for temperature optimum determination as reported previously[19].

**Enzyme secretion —** To compare the production and processing of *Pm*Sbt and *Pm*Sbtdel, 500 mL cultures were grown as described above. Time-samples of medium and cells were taken and analyzed for enzymatic activity using an amidolytic assay, as described above.

**Enzyme identification —** To confirm the identity of isolated recombinant protease, a protein band of 40 kDa was excised from a 12% SDS gel and tryptic digestion was performed according to the manual InGelBlue (GBiosciences, Saint Louis, MO, USA). The samples were analyzed by LC-MS/MS. Peptide identification was carried out using the program Scaffold (Proteome Software, Portland, OR, USA) searching against the *E. coli* protein database to which the sequences of *Pm*Sbt and trypsin were added. Matches were accepted if they could be assigned with a confidence of identification probability higher than 99%.

**Homology modeling —** The generation of structural models was carried out using the YASARA[20] homology model building protocol[21,22]. In order to increase the quality, the obtained models were submitted to 500 picoseconds molecular dynamic refinement simulation using the protocol describe by Krieger *et al*.[23]

**Molecular dynamics (MD) simulations —** The generated homology model of the mature sequence was used as starting point for MD simulations. All MD simulations were performed with GROMACS 4.6.7[24] using the Amber ff99SB-ILDN[25] force field. The LINCS algorithm[26] was used for bond-length constraining. The non-bonded pair list was updated every 10 fs. All other titratable amino acids were assigned their canonical state at physiological pH. In order to generate a well-equilibrated starting structure, the previously generated homology model was





placed in a dodecahedral box with walls set at least 12 Å from the protein surface. Explicit SPC[27] water molecules were added, a total of 150 mM NaCl (including neutralizing counter-ions) was added to simulate physiological salinity. Following steepest descents minimization (50,000 steps, or maximum force lower than 100 kJ mol-1 nm-2), the system was equilibrated in two steps, with position restraints applied to the protein heavy atoms throughout. The first phase involved simulating for 1000 ps under a constant volume (NVT) ensemble, thus allowing complete "soaking" of the protein. The complete system was coupled to temperature baths, and temperature was maintained at 310 K using the V-rescale coupling method[28]. After NVT equilibration, 1000 ps of constant pressure (NPT) equilibration were performed, also using weak coupling[27] to maintain pressure isotropically at 1.0 bar. Production simulations were run 6 times with randomly generated initial velocities and a run time of at least for 25 ns, of which the first 5 ns were excluded from analysis and considered extra equilibration time. The simulations were then individually analyzed, and their averaged properties are reported. Programs included in the GROMACS package as well as some in-house scripts were used to perform the analysis of the trajectories. The solvent-accessible surface area (SASA) was calculated using the method reported by Eisenhaber et al.[29]. The secondary structure was monitored with the DSSP[30] protocol. Conformational flexibility was assessed by the determination of root mean square fluctuation (RMSF) values. Molecular graphics images were produced using the UCSF Chimera package[31].

**Aggregation tendency prediction —** In order to identify protein regions with putative tendency to aggregation, we used the TANGO[32-34] algorithms. A range of 0.0 to 1.0, with increments of 0.1 units of ionic strength was simulated.

**Thermostability determination —** For thermostability measurements, the thermofluor method[35] was applied as follows. To the PMSF-inhibited subtilase (20 min incubation with 10 mM PMSF) in Hepes buffer (100 mM Hepes·NaOH, pH 7.5, with 1 mM $CaCl_2$) 5 µL of 100 x Sypro Orange dye was added in a thin-wall 96-well PCR plate (total volume 25 µl). For measuring the effect of disulfide bonds on thermostability DTT-pretreated samples were also prepared (25 min incubation at room temperature with 10 mM 1,4-dithio-D-threitol (DTT)). The plates were sealed with Optical-Quality Sealing Tape (Bio-Rad) and were heated for 10 min in a CFX 96 real-time PCR system (Bio-Rad) from 20 to 99°C with a heating rate of 0.5°C/min. Fluorescence changes were monitored with a CCD camera. The wavelengths for excitation and emission were 490 nm and 575 nm, respectively.

**Effect of salts, inhibitors and detergent on *Pm*Sbt activity —** The protease inhibitor phenylmethylsulfonyl fluoride (PMSF), EDTA, and iodoacetate were tested at concentrations of 10 mM. Metal ions were used as chloride salts of calcium, magnesium, cobalt, zinc and nickel. The enzyme was preincubated with the respective reagents dissolved in 100 mM Hepes·NaOH, pH 7.5, 1 mM $CaCl_2$, 20% (*v/v*) glycerol for 1 h at 40°C, and the residual activity was determined with the amidolytic assay in duplo. The activity in the same buffer without potential inhibitor added was set at 100%.

**Effect of organic solvents on the enzyme stability —** To 200 µL of enzyme in buffer (100 mM Hepes·NaOH, 1 mM $CaCl_2$, 20% glycerol, pH 7.5) an equal volume of organic solvent was added. The reaction mixtures were incubated in glass HPLC vials at 30°C with 90 rpm shaking for up to 14 days. The residual amidolytic activity was determined as described above using 1 mM of the substrate *N*-succ-AAPF-*p*NA. In two-phase systems, carefully withdrawn water phase were used. Stability was expressed as the residual activity compared to the control (enzyme in the buffer).

**Enzymatic peptide synthesis —** Enzymatic synthesis of the hexapeptide Ac-Asp-Phe-Ser-Lys-Leu-Leu-$NH_2$ was carried out as follows. Substrate stock solutions were prepared in DMSO, i.e. 40 mM stock of acyl donors (Ac-Asp-Phe-Ser-Lys-Leu-OCam and Ac-Asp-Phe-Ser-Lys-Leu-OH) (OCam stands for the glycolamide ester) and 160 mM nucleophile (H-Leu-$NH_2$) stock. The substrates and if required an internal standard (*p*-toluenesulfonic acid, 40 mM in DMSO) were added to the 1M Hepes buffer. pH 8.0. Final concentrations were 20 mM acyl donor, 80 mM nucleophile, 1 mM DTT, 1 mM internal standard (optional) and 50% (*v/v*) DMSO in 1M Hepes·NaOH buffer, pH 8.0. The presence of DMSO was necessary to solubilize the acyl donor. Reactions were started by addition of enzyme (11 nM) and the mixtures were shaken for 48 h at 150 rpm at 30°C. Samples (25 µL) were taken at different times, quenched (75 µL of acetonitrile-water (3:1, *v/v*), 1% formic acid) and analyzed with HPLC and LC-MS.





**HPLC and LC-MS —** Peptide coupling reactions were analyzed by HPLC as reported in Chapter 5 with one difference. Elution was done with a gradient as follows: 0-23 min linear gradient from 10%-30% eluent B in A, continued for 7 min to 50% eluent B, then 10 min at 10% eluent B. Separation was carried out at 25°C, with detection at 220 nm. The identity of synthetic products was confirmed by LC-MS as reported in Chapter 5.

## Results and Discussion

A novel putative protease gene was discovered by searching the genome of *Pseudomonas mendocina* ymp for sequences related to subtilisin E[19]. This putative protease (YP_001189588) shows 38% sequence similarity to the query sequence and was termed *Pm*Sbt. The encoded protein belongs to the subtilisin subfamily[19] according to Siezen's classification36. Comparison to other subtilases suggests the presence of an N-terminal domain (encompassing residues Met1-Met127, Figure 1) and a relatively long C-terminal region (Ala128-Glu520) including the catalytic core with a catalytic triad (Ser350, His193, Asp157) (Figure 1).

```
Proteinase K      ------MRLS--VLL--S--LLPLALGAPAVEQRSEAAPLI------------EARGEM   35
Aqualysin         ------MRKT--YWLMAL--FAVLVLGGCQMASRSDPTPTLAEAFWPKEAPVYGLDDPEA   50
proN-Tk-SP        APQKPAVRNV-----SQQKNYGLLTPGLFKKVQRM---------------------SWD   33
Proteolysin       ------MKKI---LLTLV--IAVLLLSGFAGVKSAEL--LF--------VSNSTTTNQED   39
PmSbt             ------MQGFQRT----T-VLLLLLSAGWTPVQADEK-PSF--------AAQAQIAVDDQ   40
Thermitase        ------MKNKIIVFLSVL---SFIIGGFFF----N----TN--------TSSAETSSTDY   35
Carlsberg         -----MMRKK-SFWLGMLTAFMLVFTMAF----------SD--------SASAAQ-PAKN   35
Subtilisin E      ------MRSK-KLWISLLFALTLIFTMAF----------SN--------M-SVQAAGKSS   34
BPN'              ------MRGK-KVWISLLFALALIFTMAF----------GS--------TSSAQAAGKSN   35
                                                                    ::           :

Proteinase K      VANKYIVKFKEGSALSALDAAME---------K--ISG-------KPDHVYKNVFSGFAA   77
Aqualysin         IPGRYIVVFKKGKGQSLLQGGIT---------T--LQARLAPQGVVVTQAYTGALQGFAA   99
proN-Tk-SP        QEVSTIIMFDNQA-----------------DKEKAVEILDFLGAKIKYNY-HIIPALAV   74
Proteolysin       PENEIIVGYKENTDVAVLS---------KQVEK-TTGAKLSR---KGLKNF-AVFKLPQG   85
PmSbt             QSPRYIIKYKELAPSPMNQANQPQPLSAGRFESRAAQRLLSQAQVQPLMHL-DSQAASVA   99
Thermitase        VPNQLIVKFKQNASLSN----------VQSFHKSVGATVLSK---DDKLGF-EVVQFSKG   81
Carlsberg         VEKDYIVGFKSGVKTA------------SV-----KKDIIKESGGKVDKQF-RIINAAKA   77
Subtilisin E      TEKKYIVGFKQTMSAMS----------SAK-----KKDVISEKGGKVQKQF-KYVNAAAA   78
BPN'              GEKKYIVGFKQTMSTMS----------AAK-----KKDVISEKGGKVQKQF-KYVDAASA   79
                                                                    *: :..
```

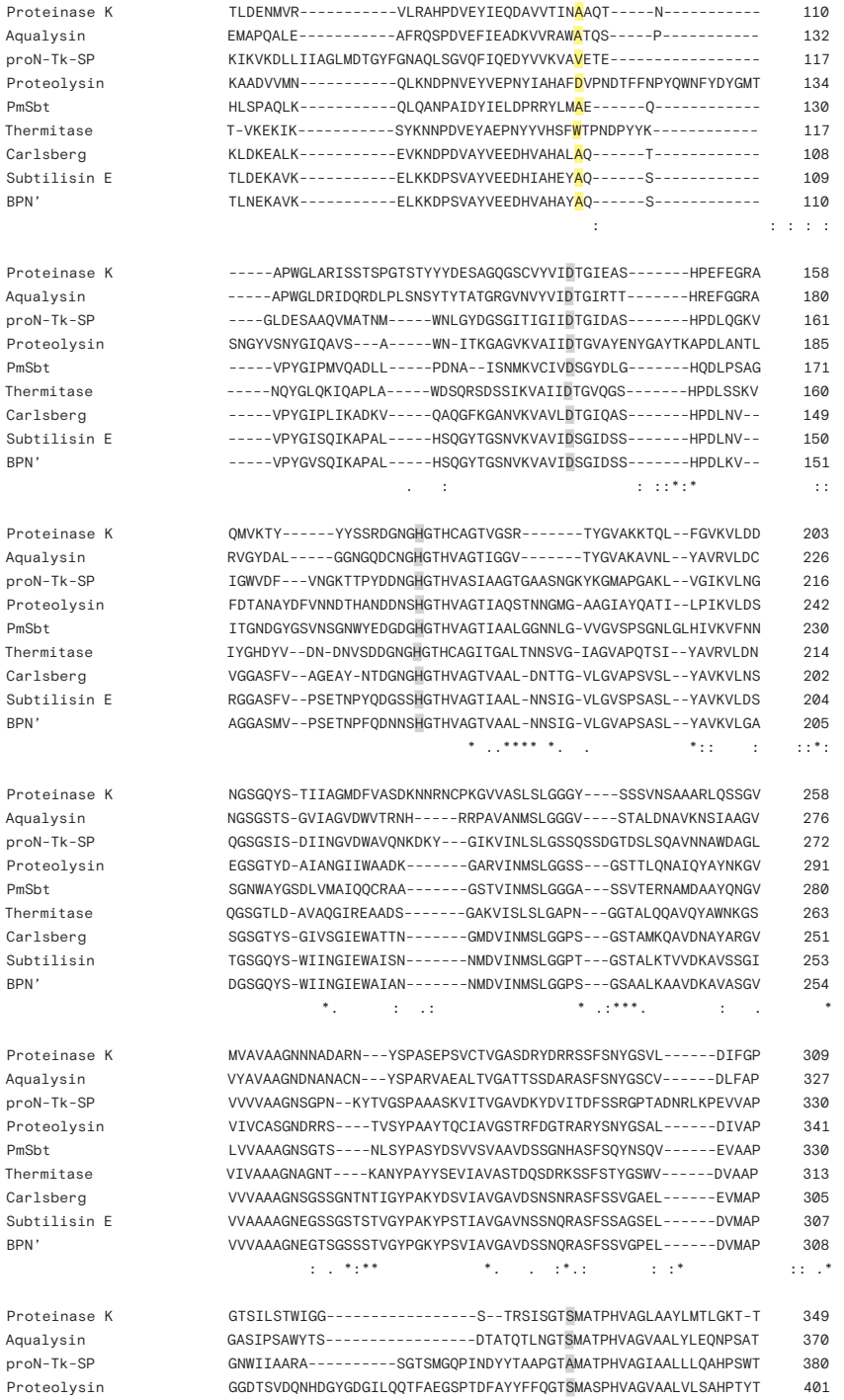



FIGURE 1.

Sequence alignment of *Pm*Sbt and other subtilases made by ClustalO software. Subtilisin E (P01489), subtilisin BPN' (P00782), subtilisin Carlsberg (P00780), thermitase (P04072), proteolysin (YP_002247839), pro-N-Tk-SP (3AFG), aqualysin I (P08594), proteinase K (P06873). The first amino acid of the mature domain is indicated in yellow and the C-terminal Arg420 of truncated *Pm*Sbtdel in cyan. If present, the first amino acid of the C-terminal prodomain is indicated in red. The active site residues are marked in grey (Ser350, His193, Asp157 in *Pm*Sbt).

In order to obtain structural information for this novel protease, a series of homology models were created using the YASARA suite and subsequently refined by MD simulations, as described in Materials and Methods (Table 1). Model 0 was generated using only proN-Tk-SP as template (3AFG), while models 1 and 2 correspond to the precursors of *Pm*Sbt and *Pm*Sbtdel, respectively, and were created using several templates. Structure of subtilisin E (PDB:1SCJ) contributed the most to the final hybrid model. Models 3 and 4 were made using predicted mature sequences of *Pm*Sbt (residues 128-520) and *Pm*Sbtdel (residues 128-420) and several templates. Here, the structure of subtilisin Carlsberg complexed with the inhibitor greglin (PDB: 4GI3) was the main contributor to the final hybrid

model. The homology models provided reliable structural information for residues Ala33 to Arg420 of the full sequence of *Pm*Sbt (model 3), confirming the presence of signal sequence followed by a central domain with a typical subtilase fold structure. However, there were no suitable templates to generate a high quality model of the C-terminal domain which was predicted in model 0 to have a $\beta$-jelly roll structure, similar to what is observed in *Tk*-SP from *T. kodakaraensis* (PDB: 3AFG)[37]. This C-terminal domain also shares some structural similarities to that of an unpublished domain that is part of a metalloprotease (PDB: 2LUW). A homology model obtained with the crystal structure 3AFG as template was of low quality (model 0) but still illustrates the domain organization of *Pm*Sbt (Figure 2).

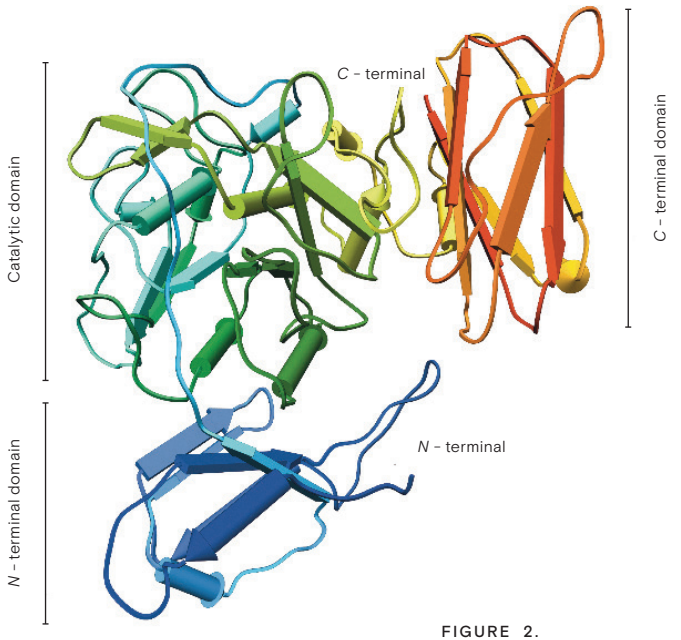



**FIGURE 2.**

Domain organization of the *Pm*Sbt. This 3D model was generated based on a single structural template (3AFG). The N-terminal domain (blue) is removed by autocatalytic processing. The C-terminal domain has a predicted $\beta$-jellyroll structure but the function is unknown.



| Model no. | Enzyme | Sequence range | Structural information range obtained | Template | Number of models | Z-score hybrid model | Z-score after MD |
|---|---|---|---|---|---|---|---|
| 0 | *Pm*Sbt | 1-520 | V21-E520 | 3AFG (Tk-SP) | 5 | -2.525 (Poor) | -2.25 (Poor) |
| 1 | *Pm*Sbt | 1-520 | A33-A416 | 1SCJ (subtilisin E) | 21 | -1.75 (Satisfactory) | -1.56 (Satisfactory) |
| 2 | *Pm*Sbtdel | 1-420 | A33-A416 | 1SCJ (subtilisin E) | 22 | -1.69 (Satisfactory) | -1.52 (Satisfactory) |
| 3 | ***Pm*Sbt Mature** | 128-520 | A128-N421 | 4GI3 (subtilisin Carlsberg complexed with inhibitor) | 17 | -0.96 (Good) | -0.93 (Good) |
| 4 | *Pm*Sbtdel Mature | 128-420 | A128-R420 | 4GI3 (subtilisin Carlsberg complexed with inhibitor) | 20 | -1.18 (Satisfactory | -1.10 (Satisfactory) |

**TABLE 1.**

Homology models of the *P. mendocina* proteases. The amino acid sequence range of the target subtilase variants, the selected highest contributing templates, the amino acid range with trustable structural information, and the number of models generated per run is shown. The best model was used for further structural analyses and is marked in red. Moreover, the accuracy of the generated models is reported by the use of Z-scores. The models were subjected to MD-simulations and improved Z-scores are reported.

The N-terminal prodomain (residues 33-120), that is removed upon maturation showed a typical $\alpha/\beta$ organization with connections by short loop regions. Interestingly section 121-131 showed a rather extended loop that passes across the active site of the enzyme, thus allocating residues Ala128 and Met127 within reach of the catalytic triad (Figure 3, panel A). These structural observations indicate that autocatalytic cleavage of the N-terminal is most probable to happen at the peptide bond between residue Met127 and Ala128, thus leaving residue 128 as the first residue of the mature sequence. After MD relaxation the catalytic Ser350 got into close proximity to the carboxylic group of Met127 (about 3.8 Å, distance between Ser Oγ and Met127 carbonyl carbon), thus suggesting the possibility of a nucleophilic attack at this position, suggesting cleavage of the Met127-Ala128 peptide bond.

The rest of the model (residues 131-420) shows the characteristic structural organization of subtiliases with a 7-stranded parallel β-sheet (in order 2314567,

Figure 3, panel B) interlaced by series of loops and turns with a total of 9 α-helix elements, and 2 antiparallel β-sheets towards the C-terminus. The acid member of the triad (Asp157) is located at the extreme of the β1 element, while the base (His193) and the nucleophile (Ser350) are located at the beginning of the α4 and α7 elements, respectively (Figure 3, panels B and C). The active site triad resides are in a classical configuration, in which Asp157 makes a hydrogen bond with His193, and His193 forms a hydrogen bond with Ser350. Three loops seem to encompass the bedding of the substrate binding site; β3-α5 (Phe228-Tyr236), β4-α6 (Leu259-Ser265), and β5-β6 (Ala286-Ser302) are depicted in Figure 3 (panel B and C, in green). These three loops showed high conformational flexibility; especially high RMSF values were observed for the loop connecting the structural elements β3-α5 (with RMSF values of around 2 Å), thus suggesting a great plasticity in the substrate-binding site (Figure 3). Although significant conformational flexibility is observed, the overall secondary structure of this model was conserved along the simulation time.

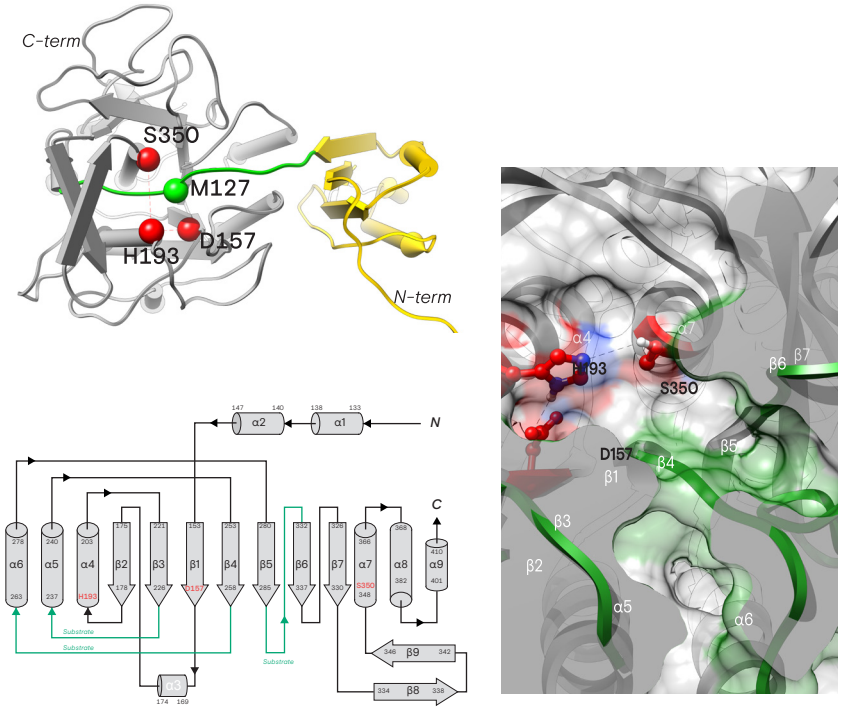






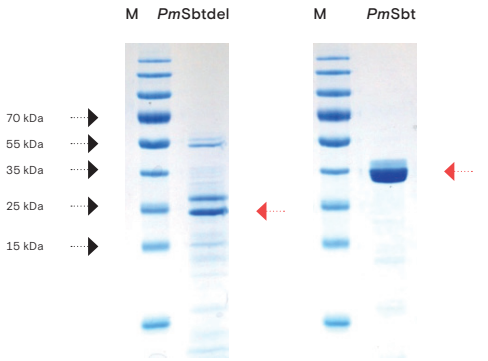

**FIGURE 3.**

Structural organization of the *Pseudomonas mendocina* protease (*Pm*Sbt). Panel A shows the homology model generated for the full-sequence of *Pm*Sbt. The N-terminal domain is depicted in yellow. Segment 121-131 which passes across the active site is in green, while the predicted cleavage point (Met127) is shown as a green sphere. The catalytic triad residues are depicted with red spheres. The mature catalytic domain is depicted in grey, while the *N*-terminal domain is shown in yellow. Panel B shows a schematic representation of the secondary structure topology of the mature *Pm*Sbt, with *α*-helices shown as cylinders, *β*- strands as block-arrows, and loops and turns as lines. This organization was found in all the homology models generated. The approximate residue numbers of the structural elements, the location of the catalytic triad residues (in red font), and the location of the loops encompassing the substrate binding site (green lines) are indicated. Panel C shows a close-up of the active site of *Pm*Sbt. The catalytic triad residues are depicted in red and the loops surrounding the substrate-binding site are shown in green.

The model structure was inspected for potentially stabilizing features. Three disulfide bonds were predicted (Cys154-Cys247, Cys370-Cys414, Cys479-Cys490). The first is in the catalytic core, the second, if present, may connect the catalytic domain and the C-terminal domain, whereas the third is predicted to be within the C-terminal domain (predicted β-jellyroll). The cysteines are missing in the C-terminal jellyroll domain of *T. kodakaraensis* subtilase Tk-SP, but on the other hand two calcium binding sites important for the Tk-SP thermostability were identified[37]. Sequence comparison reveals that four out of five residues forming calcium site 1 in Tk-SP are conserved in the *Pm*Sbt sequence (Asp461, Asp463, Leu461 and Glu488 (*Pm*Sbt numbering), but the calcium site 2 of Tk-SP37 is not clearly conserved in *Pm*Sbt. On the other hand, two calcium binding sites with subtilisin E topology (Ca-site 1: Glu129, Asp166, Leu194, Asn197, Leu199, Val201; and Ca-site 2: Ala298, Tyr300, Val303 (*Pm*Sbt numbering) are conserved in the sequence of *Pm*Sbt.

The gene coding for *Pm*Sbt was PCR amplified and recombinantly expressed in *E. coli* with a hexahistidine tag fused to the C-terminus. In order to investigate the role of the extended C-terminal region, a truncated variant called *Pm*Sbtdel in which the C-terminal Arg420 just before the proposed jellyroll domain is fused to the hexahistidine tag was also constructed. The majority of the enzyme (85%) was found in the medium of *E. coli* cells expressing *Pm*Sbt, and only a small fraction could be isolated from the cell-free extract (15%). In contrast, the proteolytic activity of the truncated variant *Pm*Sbtdel was more equally distributed over cell-free extract and medium (59% and 41%, respectively), suggesting that the C-terminal domain also facilitates enzyme secretion in the *E. coli* host. Apparently, the C-terminal domain did not block processing of the protease precursor in *E. coli* since also the truncated variants showed proteolytic activity.

The C-terminal hexahistidine tag allowed purification of both *Pm*Sbt and *Pm*Sbtdel using IMAC chromatography. When isolated from the medium, the enzyme was precipitated with ammonium sulfate, resuspended in buffer, desalted and then applied onto His Trap column. Glycerol had to be added to the buffers (20% v/v) in order to solubilize the protein pellet after ammonium sulfate precipitation. Without that, the *Pm*Sbt and *Pm*Sbtdel enzymes showed a tendency to aggregate. Using the protocol described under Materials and Methods, 10 mg of purified *Pm*Sbt and 11 mg of *Pm*Sbtdel protein were isolated from 1 L of recombinant culture. MS performed on tryptic digests of the excised 40 kDa band of pure *Pm*Sbt and the 33 kDa protein band of *Pm*Sbtdel subtilase confirmed their identity and suggested Ala128 as the first residue of the mature domain (Figure 4).




### *Pm*Sbt

12 exclusive unique peptides, 16 exclusive unique spectra, 21 total spectra, 162/520 amino acids (31% coverage)

MQGFQRTTVL LLLLSAGWTP VQADEKPSFA AQAQIAVDDQ QSPRYIIKYK ELAPSPMNQA NQPQPLSAGR FESRAAQRLL
SQAQVQPLMH LDSQAASVAH LSPAQLKQLQ ANPAIDYIEL DPRRYLMAEQ VPYGIPMVQA DLLPDNAISN MKVCIVDSGY
DLGHQDLPSA GITGNDGYGS VNSGNWYEDG DGHGTHVAGT IAALGGNNLG VVGVSPSGNL GLHIVKVFNN SGNWAYGSDL
VMAIQQCRAA GSTVINMSLG GGASSVTERN AMDAAYQNGV LVVAAAGNSG TSNLSYPASY DSVVSVAAVD SSGNHASFSQ
YNSQVEVAAP GVGVRSTLPG NRYASYNGTS MATPHVSALY ALVWSQHRQC TPAQIRRVVN ITAEDRGTPG RDPYYGYGIV
KAKRASDLIA QRGCDASDGR NGEEQTYPNL SGARGAWVRH SVAVPSGARR LTVRISGGSG DADLYTHLGS QPTTSQWTCR
PYLEGNQETC IQENPAAGTW HIGVRGYSSF SGVTLYWRYE

### *Pm*Sbtdel

9 exclusive unique peptides, 10 exclusive unique spectra, 28 total spectra, 77/520 amino acids (15% coverage)

MQGFQRTTVL LLLLSAGWTP VQADEKPSFA AQAQIAVDDQ QSPRYIIKYK ELAPSPMNQA NQPQPLSAGR FESRAAQRLL
SQAQVQPLMH LDSQAASVAH LSPAQLKQLQ ANPAIDYIEL DPRRYLMAEQ VPYGIPMVQA DLLPDNAISN MKVCIVDSGY
DLGHQDLPSA GITGNDGYGS VNSGNWYEDG DGHGTHVAGT IAALGGNNLG VVGVSPSGNL GLHIVKVFNN SGNWAYGSDL
VMAIQQCRAA GSTVINMSLG GGASSVTERN AMDAAYQNGV LVVAAAGNSG TSNLSYPASY DSVVSVAAVD SSGNHASFSQ
YNSQVEVAAP GVGVRSTLPG NRYASYNGTS MATPHVSALY ALVWSQHRQC TPAQIRRVVN ITAEDRGTPG RDPYYGYGIV
KAKRASDLIA QRGCDASDGR NGEEQTYPNL SGARGAWVRH SVAVPSGARR LTVRISGGSG DADLYTHLGS QPTTSQWTCR
PYLEGNQETC IQENPAAGTW HIGVRGYSSF SGVTLYWRYE





**FIGURE 4.**

SDS-PAGE of purified *Pm*Sbtdel and *Pm*Sbt subtilases. M, marker; PageRuler Prestained Protein Ladder (ThermoFischer). The excised bands are indicated with a red arrow. Results of tryptic digestion of *Pm*Sbt 40 kDa (PM40B) excised band and *Pm*Sbtdel 35kDa 9(R35) excised band. Identification is performed by Scaffold software. Bottom: position of identified peptide segments after trypsin digestion in the prosequences of *Pm*Sbtand *Pm*Sbtdel proteases (in yellow). Residues in green indicate amino acids for which modifications were allowed by the software.

To explore the aggregation behavior, we analyzed the sequence of PmSbt with the statistical mechanics algorithm TANGO32, which predicts aggregation-prone segments from sequence information. For *Pm*Sbt 10 sequence segments with potential aggregation properties were detected. Three segments with low aggregation tendency (lower than 5%) were found: 197-205 (L1), 209-214 (L2), and 394-400 (L3). Four segments with medium aggregation tendency (aggregation between 5% - 50%) were identified: 240-245 (M1), 302-309 (M2), 378-383 (M3), and 507-517 (M4). Three segments with high aggregation tendency (>50%) were found: 7-20 (H0), 276-288 (H1), and 355-366 (H2). Segments H1 and H2 showed aggregation tendency values of above 90% (predicted percentage of aggregation for a fully solvent-exposed unfolded peptide) (Figure 5a). Since these segments are likely responsible for the aggregation propensity we investigated their position in the structural model (Figure 5b). Interestingly, H1 encompasses the end section of α5 and the complete β5, showing a very low solvent exposure and very low RMSF values for the residues with the highest aggregation tendency. However, the flanking reside Gly279, which shows an aggregation tendency of 10.3±0.16%, an RMSF value of 0.67±0.03 Å, and a solvent exposure value of 15.9±2.1%, could work as a trigger towards aggregation under specific conditions due to its partial solvent exposure and moderate flexibility (Figure 5C). On the other hand, aggregation prone segment H2 encompasses the majority of helix α7 and shows a higher flexibility than H1. This suggests a more important role of H2 in the aggregation. Residue Leu362, which shows an aggregation tendency of 93.4±0.18%, with an RMSF value of 1.04±0.15 Å and a solvent exposure value of 12.0±3.72%, could be the main trigger for the formation of aggregates (Figure 5D).

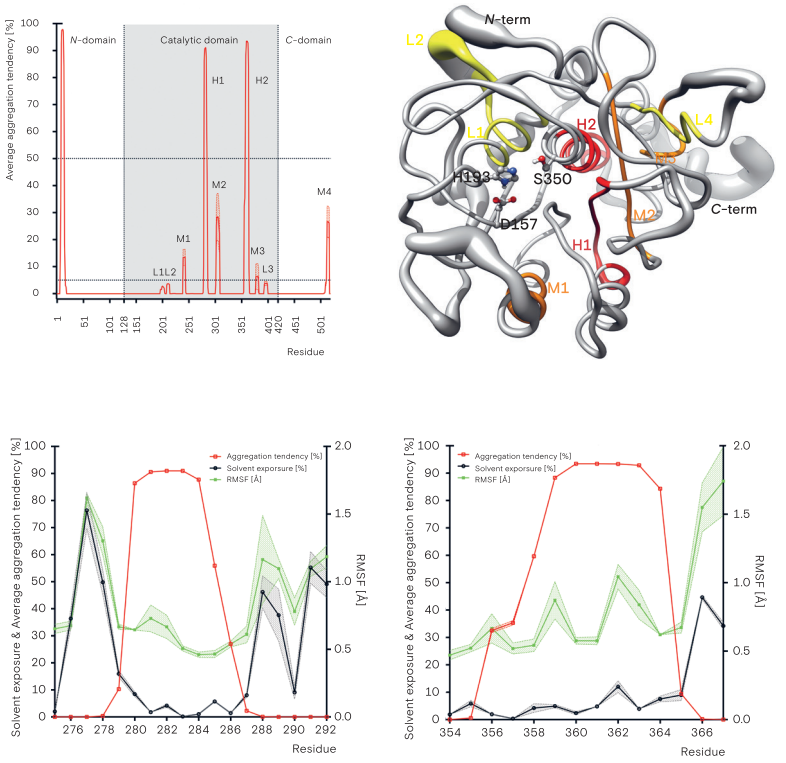



FIGURE 5.

Aggregation tendency prediction for the *P. mendocina* protease (*Pm*Sbt). Panel A shows the results obtained from the TANGO algorithm. The catalytic domain of *Pm*Sbt (residue 128-420) is depicted in a grey box. Panel B shows the 3D location of the segments with potential aggregation tendencies. Segments with low aggregation tendency are depicted in yellow, medium tendency in orange and high tendency in red. Also, the location of the catalytic triad residues is shown. Moreover, the conformational flexibility of the system as observed in MD simulations is depicted in "worm" tube representation, where large diameter tubes indicate high flexibility and vice versa. The conformational flexibility was estimated based on the RMSF calculation of each residue. Panels C and D show the aggregation tendency [%], solvent exposure [%] and RMSF [Å] for the segments H1 and H2, respectively. The error lines indicate the standard deviation upon 6 independent MD simulations.





Further biochemical characterization of the isolated enzymes indicates 46°C as optimal temperature of PmSbt with azocasein as the substrate (Figure 6). The apparent melting temperatures (Tm,app) as determined by the thermofluor method were 54.5 °C and 53.3 °C for *Pm*Sbt and PmSbtdel, respectively, implying that deletion of the C-terminal domain only slightly decreased thermostability. The optimum pH in the amidolytic assay was 8.0 for PmSbt subtilase.

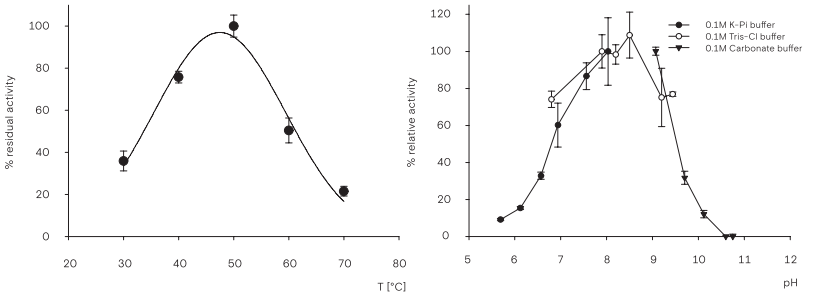



FIGURE 6.

Temperature and pH profile of *Pm*Sbt. Panel A: temperature profile determined in triplicate using the azocasein assay. Panel B: pH curve determined in triplicate using phosphate buffer (0.1 M, pH range 5.7-8.0), Tris buffer (0.1M, pH 6.8-9.4) and carbonate buffer (0.1 M, pH 9.0-10.8) in the amidolytic assay.

As mentioned earlier, many organic-solvent tolerant enzymes (lipases, esterases) and metalloproteases originate from *Pseudomonas* genera[6]. To examine the tolerance of the *Pm*Sbt protease to organic cosolvents, a set of 20 cosolvents with varying logP values was used in 50% (*v/v*) concentration (Table 2). As expected, hydrophobic solvents in a two-phase system (logP>2) did not inhibit the protease as determined in amidolytic assays. Hydrophilic solvents (logP<2) caused inactivation within 5 days (acetonitrile, THF, chloroform, alcohols). Exposure to acetone, isopropanol and dioxane caused 70% loss of activity in 5 days. Longer incubation resulted in complete loss of activity. Surprisingly, *Pm*Sbt exhibited remarkable stability in the presence of DMSO and DMF as cosolvents (50% *v/v*) for up to 14 days. Also dichloromethane appeared well tolerated. A similar behavior was observed with *Pm*Sbt and the truncated enzyme *Pm*Sbtdel when activity was assayed using azocasein as the substrate (data not shown).

| logP | Solvent | Solvent type | Solvent system | Residual activity | |
|---|---|---|---|---|---|
| | | | | day 5 | day 14 |
| -1.412 | dimethylsulfoxide | polar aprotic | cosolvent | 0.81 | 0.92 |
| -1.362 | ethane-1,2-diol | polar protic | two-phase | ND | 0.55 |
| -0.829 | *N.N*-dimethyl formamide | polar aprotic | cosolvent | 0.52 | 0.47 |
| -0.69 | methanol | polar protic | cosolvent | 0.07 | 0.01 |
| -0.334 | acetonitrile | polar aprotic | cosolvent | 0.00 | 0.00 |
| -0.255 | 1,4-dioxane | nonpolar | cosolvent | 0.25 | 0.06 |
| -0.18 | ethanol | polar protic | cosolvent | 0.08 | 0.00 |
| -0.042 | acetone | polar aprotic | cosolvent | 0.24 | 0.05 |
| 0.173 | propan-2-ol | polar protic | cosolvent | 0.27 | 0.05 |
| 0.329 | propanol | polar protic | cosolvent | 0.00 | 0.00 |
| 0.473 | tetrahydrofuran | polar aprotic | cosolvent | 0.00 | 0.00 |
| 1.405 | dichloromethane | slightly polar | two-phase | 0.43 | 0.39 |
| 1.935 | chloroform | slightly polar | two-phase | 0.01 | 0.00 |
| 2.177 | benzene | nonpolar | two-phase | 0.67 | 0.54 |
| 2.72 | toluene | nonpolar | two-phase | 0.70 | 0.62 |
| 3.163 | cyclohexane | nonpolar | two-phase | 0.68 | 0.62 |
| 3.764 | hexane | nonpolar | two-phase | 0.64 | 0.62 |
| 4.783 | octane | nonpolar | two-phase | 0.71 | 0.62 |
| 5.802 | decane | nonpolar | two-phase | 0.68 | 0.22 |
| 6.821 | dodecane | nonpolar | two-phase | 0.78 | 0.62 |
| | buffer | none | none | 0.91 | 0.81 |

**TABLE 2.**

Effect of organic solvents on *Pm*Sbt stability. The enzyme was incubated with 50% (*v/v*) cosolvent and 10% (*v/v*) glycerol. The stability was determined as the remaining activity relative to the control (enzyme in buffer) using the amidolytic assay as described in Materials and Methods. Standard deviation was less than 9% except in case of dichloromethane after 14 days (14%).





The organic solvent tolerance could make the *Pm*Sbt enzyme a valuable protease when substrate solubility needs to be increased by cosolvent addition. In order to compare the *Pm*Sbt enzyme with other subtilases, the steady-state kinetic parameters of *Pm*Sbt and *Pm*Sbtdel were determined (Table 3) using the standard amidolytic substrate. The *K*m for this substrate is high. With addition of thiol-reducing agent DTT (10 mM), both the *K*m and the kcat of *Pm*Sbt for *N*-suc-AAPF-*p*NA increased, resulting in lower overall enzyme efficiency as compared to the reaction without DTT addition. In case of *Pm*Sbtdel with the truncated *C*-terminus, the enzyme is saturated with substrate at lower concentrations. Under these conditions subtilases *Pm*Sbt and *Pm*Sbtdel catalyzed hydrolysis of *N*-suc-AAPF-*p*NA with far lower efficiency than subtilisin A.

| Enzyme | -DTT | | | +DTT | | | reference |
|---|---|---|---|---|---|---|---|
| | $k_{cat}$ [$s^{-1}$] | $K_m$ [mM] | $k_{cat}/K_m$ [$mM^{-1}\ s^{-1}$] | $k_{cat}$ [$s^{-1}$] | $K_m$ [mM] | $k_{cat}/K_m$ [$mM^{-1}\ s^{-1}$] | |
| *Pm*Sbt | 10.5±1.4 | 2.98±0.73 | 3.5 | 19.8±4.3 | 7.4±2.2 | 2.7 | this work |
| *Pm*Sbtdel | 8.4±3 | 0.99±0.24 | 8.5 | ND | ND | ND | this work |
| Subtilisin A | 551±62 | 1.1±0.3 | 489 | 40±3 | 1.57±0.25 | 23 | [19] |

**TABLE 3.**

Kinetic parameters of proteases. All reactions were performed in 100 mM Hepes·NaOH buffer, pH 7.5, containing 1 mM $CaCl_2$ and 10% DMSO (*v/v*) at 40°C using *N*-suc-AAPF-*p*NA as the substrate. Where indicated the reaction was supplemented with 10 mM DTT.

Next, the effects of various metal ions, protease inhibitors and detergents on *Pm*Sbt activity were determined by measuring residual amidolytic activity after 1 h preincubation under appropriate conditions (Table 4). The serine protease inhibitor PMSF abolished the activity of protease, confirming that *Pm*Sbt belongs to the serine protease class. In addition,

the metal chelator EDTA strongly reduced *Pm*Sbt activity suggesting a crucial role of the predicted metal ion binding site (see text above) for enzyme integrity or activity. However, not every bivalent cation was accepted; by addition of $Ca^{2+}$ ions the activity was doubled, as observed for proteolysin (Chapter 4). On contrary, the addition of $Ni^{2+}$ and $Co^{2+}$ proved to be deleterious for the enzymatic activity.

| Salts, inhibitors, detergents | Concentration | % Residual activity |
|---|---|---|
| | | *Pm*Sbt |
| None | - | 100 |
| $Ca^{2+}$ | 10 mM | 218±6 |
| $Ni^{2+}$ | 10 mM | 1.6±2.1 |
| $Co^{2+}$ | 10 mM | 1.4±1.0 |
| SDS (sodium dodecyl sulfate) | 10% (*w/v*) | 0.1±0.1 |
| cetyl trimethylammonium chloride | 12.5% (*w/v*) | 0.4±0.2 |
| Tween 20 | 10% (*w/v*) | 281±11 |
| PMSF | 10 mM | 0.1±0.1 |
| EDTA | 10 mM | 1.5±0.2 |
| iodoacetamide | 10 mM | 170±2 |
| DTT | 10 mM | 137±5 |
| urea | 6 M | 0.2±0.3 |
| guanidinium chloride | 6 M | 0.5±0.6 |

**TABLE 4.**

Effect of different agents on *Pm*Sbt activity. Enzymes were incubated for 1 h at 40°C in 100 mM Hepes·NaOH buffer, pH 7.5, supplemented with the agent indicated. Remaining activities were measured in triplicate with *N*-suc-AAPF-*p*NA as the substrate..





Another activator was the nonionic detergent Tween 20 (2.8-fold increase in activity). The ionic detergents and high concentrations of urea or guanidinium chloride (6M) were not tolerated by the *Pm*Sbt subtilase. Iodoacetamide, an irreversible inhibitor of cysteine peptidases, did not have an inhibitory role; in contrast, it exhibited a positive effect on activity as did DTT (dithiothreithol) which disrupts disulfide bonds. The effect of DTT and iodoacetamide on the activity was higher for *Pm*Sbt than for the other subtilases reported in this thesis which prompted us to examine sulfhydryl groups and potential disulfide bonds.

As described above, inspection of the sequence and the structural model identified six cysteines that may form disulfide bonds in *Pm*Sbt subtilase. *Pm*Sbt CysxAla variants (x=154,247,370,414,479,490) were tested for functional expression using a skim milk plate assay. The activity of the variants was further quantified in a culture medium using amidolytic substrate *N*-succ-AAPF-*p*NA (Table 5).

| Mutation | Disrupted disulfide bond suggested by the model | Proteolytic activity on 1% skim milk plate | Proteolytic activity in liquid culture, U/mL | |
|---|---|---|---|---|
| | | | medium | CFE extract |
| Cys154Ala | Cys154-Cys247 | - | NA | NA |
| Cys247Ala | | – | NA | NA |
| Cys370Ala | Cys370-Cys414 | – | NA | NA |
| Cys414Ala | | – | NA | NA |
| Cys479Ala | Cys479-Cys490 | + | 1.2 | 0.4 |
| Cys490Ala | | + | 0.5 | 0.3 |

**TABLE 5.**

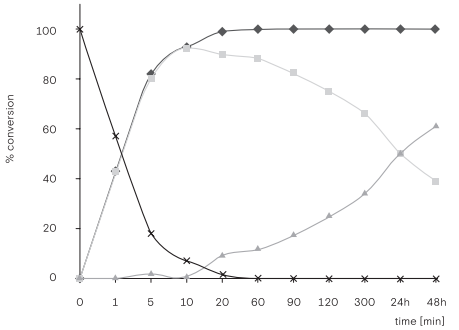
Alanine scan of putative disulfide bond forming cysteines. Proteolytic activities of the mutants was tested by plating on a skim-milk plates, and in liquid culture using the amidolytic assay described in Materials and Methods.

By disrupting the Cys154-Cys247 and Cys370-Cys414 bonds in the catalytic domain, the functional expression of the PmSbt subtilase was impaired. On the other hand, the Cys479-Cys490 bond predicted in the *C*-terminal *β*-jellyroll domain could be disrupted without deleterious effect on enzyme expression and activity.

One of the applications of subtilases is chemoenzymatic peptide synthesis. The subtilase *Pm*Sbt was tested in the synthesis of the hexameric segment of the bioactive human peptide PHM27(9-14), a potent agonist of the human calcitonin receptor[38]. Enzymatic synthesis of the hexapeptide Ac-Asp-Phe-Ser-Lys-Leu-Leu-$NH_2$ was attempted out by mixing activated acyl donor Ac-Asp-Phe-Ser-Lys-Leu-OCam and nucleophile H-Leu-NH2, both dissolved in 50% (v/v) DMSO in buffer of pH 8. Samples were analyzed by LC-MS, which revealed formation of hydrolytic product Ac-Asp-Phe-Ser-Lys-Leu-OH (retention time 16.3 min, m/z (ESI+)=651.7, $[M+H]^+$ calcd. for $C_{30}H_{47}N_6O_{10}+$ = 651.72). Surprisingly, this hydrolysis was was followed by peptide synthesis as shown by the formation of synthetic product, which was identified by LC-MS as the expected Ac-Asp-Phe-Ser-Lys-Leu-Leu-$NH_2$ (retention time 18.8 min, m/z (ESI+)=763.9, $[M+H]^+$ calcd. for $C_{36}H_{59}N_8O_{10}+$ = 763.89). This indicates that thermodynamically controlled peptide coupling occurred after the initial hydrolysis reaction, which was kinetically preferred: ester hydrolysis occurred fast and initially within 20 min all ester was hydrolyzed to acid (Figure 7). After 48 h, in the presence of 50% DMSO, 60% of the acyl donor was converted to synthetic product. Thus, under these conditions the overall equilibrium favored coupling product over free peptide segments, and *Pm*Sbt efficiently catalyzed the coupling reaction.

FIGURE 7.

Chemoenzymatic synthesis of the bioactive human peptide PHM27 (9-14) segment by *Pm*Sbt subtilase. The reaction was performed in a 50% (*v/v*) DMSO-buffer mixture at 30°C using a Cam ester as activated acyl donor (Ac-Asp-Phe-Ser-Lys-Leu-OCam, 20 mM) and H-Leu-$NH_2$ (80 mM) as the nucleophile. Starting OCam ester (x), hydrolytic byproduct (Ac-Asp-Phe-Ser-Lys-Leu-OH, squares), product (Ac-Asp-Phe-Ser-Lys-Leu-Leu-$NH_2$, triangles) and conversion (diamonds) were followed by HPLC analysis of time course samples.








Peptide synthesis under thermodynamic control was further tested using Ac-Asp-Phe-Ser-Lys-Leu-OH as acyl donor and leucine amide as the nucleophile. The synthesis was performed both in the presence of 50% (*v/v*) DMSO and with 50% DMF as the cosolvents and samples were analyzed by LC-MS After 48 h 67% conversion to the coupling product mentioned above was achieved by *Pm*Sbt in both cases. Peptide-coupling under thermodynamic control was reported also for the serine protease trypsin in reactions containing 50% DMF (*v/v*), producing α-melanocyte stimulating hormone in an 8+5 coupling in 28% yield after 48 h[39].

Often, in thermodynamically controlled enzymatic peptide couplings with metalloproteases (i.e. thermolysin) the product is precipitated in order to shift the equilibrium towards synthesis. Here, we could shift the equilibrium to the synthetic product (67% conversion) by adding cosolvent in 50% (*v/v*) ratio, without the need for product precipitation. However, in reactions with prolonged reaction times (> 24 h) a byproduct was observed close to the synthetic product in the HPLC elution profile. Mass spectrometry analysis suggested both correspond to the expected synthetic product; one component may be C-terminally deamidated coupled peptide.

Although chemoenzymatic peptide synthesis under kinetic control is faster and gives higher yields than thermodynamically controlled peptide synthesis, it requires activated acyl donors that need to be synthesized solid- or solution-phase chemical synthesis to attach the activating and protecting groups. In contrast, peptides used in thermodynamic coupling do not require an activating C-terminal group and can be produced otherwise (e.g. by fermentation). In that case, the N-terminal protective group can be enzymatically incorporated (acylase, peptide deformylase). As shown here, by using enzyme that tolerate high amounts of cosolvents, the reaction equilibrium can be shifted toward the synthetic product side.

In conclusion, we report the biochemical characterization of a subtilisin E homologue from *Pseudomonas mendocina* strain ymp. The enzyme was produced in *E. coli* and isolated from the culture medium by His-tag affinity chromatography. The purified subtilase was remarkably stable in the presence of cosolvents such as DMF and DMSO present at levels of up to 50% (*v/v*). Compared to subtilisin E, *Pm*Sbt, carries an additional C-terminal domain (predicted β-jellyroll), which appeared of minor importance for stability, unlike in the case of the protease from *T. kodakaraensis*[37]. Deletion of the C-terminal domain impaired secretion to the medium, but did not significantly affect activity. Homology modeling suggested the presence of 3 disulfide bonds (Cys154-Cys247, Cys370-Cys414, Cys479-Cys490), of which only Cys479-Cys490 in the C-terminal domain could

be removed by Cys→Ala mutagenesis without disturbing enzyme production. The enzyme exhibited a tendency to aggregate. The predicted aggregation tendency of regions H1 (276-288) and H2 (355-366) may contributed to this and can therefore serve as a target for mutagenesis aimed at reducing *Pm*Sbt aggregation.

In chemoenzymatic peptide synthesis the addition of cosolvents such as DMSO or DMF can increase substrate solubility and shift the reaction equilibrium towards peptide synthesis, allowing thermodynamically-controlled synthesis if the coupling enzymes remain sufficiently active[40]. The robustness of *Pm*Sbt made it possible to use this enzyme for peptide synthesis in the presence of 50% cosolvent. When the activated pentapeptide (Ac-Asp-Phe-Ser-Lys-Leu-OCam, 20 mM) was used as the acyl donor and Leu $NH_2$ (80 mM) as the nucleophile in the presence of 50% DMSO or DMF, the enzyme showed a strong kinetic preference for water as nucleophile, resulting in swift hydrolysis of the activated acyl donor. However, due to the equilibrium of the overall reaction, prolonged incubation led to formation of coupled product. The same coupling product was indeed obtained in reactions of the unactivated pentameric acyl donor and Leu $NH_2$ as the nucleophile reaching conversion of 67% after 48 h. These initial results suggest that the *P. mendocina* protease may be useful for coupling non-activated peptides by thermodynamic control. The industrial applicability will depend on substrate scope and possibilities to optimize enzyme production.

**Acknowledgments**

This project is part of the Integration of Biosynthesis and Organic Synthesis program (IBOS-2; project 053.63.014), funded by The Netherlands Organization for Scientific Research (NWO) and Advanced Chemical Technologies for Sustainability (ACTS).

**Author contribution**

AT performed experimental work, MFM performed bioinformatic and molecular modeling studies, MO performed MS analysis, TN synthesized the substrates, PJLMQ and DBJ supervised the work. AT, MFM and DBJ wrote the paper.





# References

01 — Thundimadathil, J. Cancer treatment using peptides: current therapies and future prospects. *J. Amino Acids* **2012**, 967347 (2012).

02 — Thayer, A. Improving peptides. *Chem. Eng. News* **89**, 13–20 (2011).

03 — Thayer, A. M. Making peptides at large scale. *Chem. Eng. News* **89**, 9–12 (2011).

04 — Lax, R. The future of peptide development in the pharmaceutical industry. *PharManufacturing Int. Pept. Rev.* 10–15 (2010).

05 — Mine, Y., Li-Chan, E. & Jiang, B. *Bioactive proteins and peptides as functional foods and nutraceuticals*. (eds. Mine,Y., Li-Chan,E. & Jiang, B.),Wiley-Blackwell, Oxford, UK, (2010).

06 — Doukyu, N. & Ogino, H. Organic solvent-tolerant enzymes. *Biochem. Eng. J.* **48**, 270–282 (2010).

07 — Wong, C.-H., Chen, S., Hennen, W. J., Bibbs, J. A., Wang, Y., Liu, J. L., Pantoliano, M. W., Whitlow, M. & Bryan, P. N. Enzymes in Organic Synthesis : Use of subtilisin and a highly stable mutant derived from multiple site-specific mutations. *J. Am. Chem. Soc.* **112**, 945–953 (1990).

08 — Economou, C., Chen, K. & Arnold, F. H. Random mutagenesis to enhance the activity of subtilisin in organic solvents: Characterization of Q103R subtilisin E. *Biotechnol. Bioeng.* **39**, 658–662 (1992).

09 — Ogino, H. & Ishikawa, H. Enzymes which are stable in the presence of organic solvents. *J. Biosci. Bioeng.* **91**, 109–16 (2001).

10 — Sardessai, Y. N. & Bhosle, S. Industrial potential of organic solvent tolerant bacteria. *Biotechnol. Prog.* **20**, 655–660 (2004).

11 — Gupta, A. & Khare, S. K. Enzymes from solvent-tolerant microbes: useful biocatalysts for non-aqueous enzymology. *Crit. Rev. Biotechnol.* **29**, 44–54 (2009).

12 — Ogino, H., Yasui, K., Shiotani, T., Ishihara, T. & Ishikawa, H. Organic solvent-tolerant bacterium which secretes an organic solvent-stable proteolytic enzyme. *Appl. Environ. Microbiol.* **61**, 4258–62 (1995).

13 — Xu, J., Jiang, M., Sun, H. & He, B. An organic solvent-stable protease from organic solvent-tolerant *Bacillus cereus* WQ9-2: Purification, biochemical properties, and potential application in peptide synthesis. *Bioresour. Technol.* **101**, 7991–7994 (2010).

14 — Tang, X. Y., Pan, Y., Li, S. & He, B. F. Screening and isolation of an organic solvent-tolerant bacterium for high-yield production of organic solvent-stable protease. *Bioresour. Technol.* **99**, 7388–7392 (2008).

15 — Schellenberger, V. & Jakubke, H.-D. Protease-catalyzed kinetically controlled peptide synthesis. *Angew. Chemie Int. Ed. English* **30**, 1437–1449 (1991).

16 — Awaya, J. D. & Dubois, J. L. Identification, isolation, and analysis of a gene cluster involved in iron acquisition by *Pseudomonas mendocina* ymp. *Biometals* **21**, 353–66 (2008).

17 — de Beer, R. J. A. C., Nuijens, T., Wiermans, L., Quaedflieg, P. J. L. M. & Rutjes, F. P. J. T. Improving the carboxyamidomethyl ester for subtilisin A-catalysed peptide synthesis. *Org. Biomol. Chem.* **10**, 6767–75 (2012).

18 — Toplak, A., Nuijens, T., Quaedflieg, P. J. L. M., Wu, B. & Janssen, D. B. Peptide synthesis in neat organic solvents with novel thermostable proteases. *Enzyme Microb. Technol.* **73-74**, 20–28 (2015).

19 — Toplak, A., Wu, B., Fusetti, F., Quaedflieg, P. J. L. M. & Janssen, D. B. Proteolysin, a novel highly thermostable and cosolvent-compatible protease from the thermophilic bacterium *Coprothermobacter proteolyticus*. *Appl. Environ. Microbiol.* **79**, 5625–32 (2013).

20 — Krieger, E., Koraimann, G. & Vriend, G. Increasing the precision of comparative models with YASARA NOVA - A self-parameterizing force field. *Proteins Struct. Funct. Genet.* **47**, 393–402 (2002).

21 — Krieger, E., Nabuurs, S. B. & Vriend, G. in *Structural Bioinformatics*, Vol 44 (eds. Bourne, P. E. & and Weissig, H.) 509–523 (John Wiley & Sons, Inc., Hoboken, New Jersey, USA, 2005).

22 — Venselaar, H., Joosten, R. P., Vroling, B., Baakman, C. A. B., Hekkelman, M. L., Krieger, E. & Vriend, G. Homology modelling and spectroscopy, a never-ending love story. *Eur. Biophys. J.* **39**, 551–563 (2010).

23 — Krieger, E., Darden, T., Nabuurs, S. B., Finkelstein, A. & Vriend, G. Making optimal use of empirical energy functions: force-field parameterization in crystal space. *Proteins* **57**, 678–83 (2004).

24 — Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **4**, 435–447 (2008).

25 — Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O. & Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins.* **78**, 1950–1958 (2010).

26 — Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).

27 — Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. J. Chem. Phys. **81**, 3684–3690 (1984).







28 — Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 1–7 (2007).

29 — Eisenhaber, F., Lijnzaad, P., Argos, P., Sander, C. & Scharf, M. The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *J. Comput. Chem.* **16**, 273–284 (1995).

30 — Joosten, R. P., te Beek, T. A. H., Krieger, E., Hekkelman, M. L., Hooft, R. W. W., Schneider, R., Sander, C. & Vriend, G. A series of PDB related databases for everyday needs. *Nucleic Acids Res.* **39**, D411–9 (2011).

31 — Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

32 — Fernandez-Escamilla, A.-M., Rousseau, F., Schymkowitz, J. & Serrano, L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.* **22**, 1302–1306 (2004).

33 — Linding, R., Schymkowitz, J., Rousseau, F., Diella, F. & Serrano, L. A comparative study of the relationship between protein structure and beta-aggregation in globular and intrinsically disordered proteins. *J. Mol. Biol.* **342**, 345–53 (2004).

34 — Rousseau, F., Schymkowitz, J. W. H. & Serrano, L. Protein aggregation and amyloidosis: confusion of the kinds? *Curr. Opin. Struct. Biol.* **16**, 118–126 (2006).

35 — Ericsson, U. B., Hallberg, B. M., Detitta, G. T., Dekker, N. & Nordlund, P. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal. Biochem.* **357**, 289–98 (2006).

36 — Siezen, R. J. & Leunissen, J. A. Subtilases: the superfamily of subtilisin-like serine proteases. *Protein Sci.* **6**, 501–23 (1997).

37 — Foophow, T. Tanaka, S. I., Angkawidjaja, C., Koga, Y., Takano, K. & Kanaya, S. Crystal structure of a subtilisin homologue , Tk-SP, from *Thermococcus kodakaraensis*: Requirement of a C-terminal β-Jelly roll domain for hyperstability. *J. Mol. Biol.* **400**, 865–877 (2010).

38 — Ma, J.-N., Currier, E. A., Essex, A., Feddock, M., Spalding, T. A., Nash, N. R., Brann, M. R. & Burstein, E. S. Discovery of novel peptide/receptor interactions: identification of PHM-27 as a potent agonist of the human calcitonin receptor. *Biochem. Pharmacol.* **67**, 1279–84 (2004).

39 — Nishino, N., Xu, M., Mihara, H. & Fujimoto, T. Use of hexafluoroisopropyl alcohol in tryptic condensation for partially protected precursor of α-melanocyte stimulating hormone. *Tetrahedron Lett.* **33**, 3137–3140 (1992).

40 — Guzman, F., Barberis, S. & Illanes, A. Peptide synthesis: chemical or enzymatic. *Electron. J. Biotechnol.* **10**, 279–314 (2007).

CHAPTER — 07

# Peptiligase, an enzyme for efficient chemoenzymatic peptide synthesis and cyclization in water

Ana Toplak[a,b], Timo Nuijens[b], Peter J. L. M. Quaedflieg[c], Bian Wu[a] and Dick B. Janssen[a]

[a] Biochemical Laboratory, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, 9747 AG Groningen, the Netherlands
[b] EnzyPep B.V., Urmonderbaan 22,6167 RD Geleen, the Netherlands
[c] DSM Innovative Synthesis, 6167 RD Geleen, the Netherlands

Part of this chapter has been published in: Toplak A, Nuijens T, Quaedflieg PJML, Wu B, Janssen DB. Peptiligase, an enzyme for efficient chemoenzymatic peptide synthesis and cyclization in water. Adv Synth Catal (2016) doi: 10.1002/adsc.201600017

Patent filed under: PCT/NL2015/0507011 on 9 Oct. 2015

## Abstract

We describe a novel, organic cosolvent-stable and cation-independent engineered enzyme for peptide coupling reactions. The enzyme is a variant of a stable calcium-independent mutant of subtilisin BPN', with the catalytic Ser212 mutated to Cys and Pro216 converted to Ala. The enzyme, called peptiligase, catalyzes exceptionally efficient peptide coupling in water with a surprisingly high synthesis over hydrolysis (S/H) ratio. The S/H ratio of the peptide ligation reaction is correlated to the length of the peptide substrate and proved to be >100 for the synthesis of a 13-mer peptide, which corresponds to >99% conversion to the ligated peptide product and <1% hydrolytic side-reactions. Furthermore, peptiligase does not require a particular recognition motif resulting in a broadly applicable and traceless peptide ligation technology.



## Introduction

Despite the high demand for peptides, the large scale manufacturing of peptides remains a challenging target. Chemoenzymatic peptide synthesis (CEPS), wherein chemically synthesized peptide segments are coupled together, is potentially one of the most cost-efficient technologies for the synthesis of medium-sized and long peptides (20-50 amino acids)[1]. The peptide segments can be made in high purity via conventional (solid phase) methods and condensed by enzymatic coupling to obtain the final product of interest.

Complete solid phase synthesis of long peptides leads to accumulation of by-products and cumbersome purification. Furthermore, chemical segment condensation leads to racemization and requires fully protected peptides, which are difficult to handle due to their hydrophobicity. As compared to chemical coupling, CEPS owns several notable advantages, i.e., the protection of the side-chain functionalities is not necessary, racemization is absent, and it is more environment friendly. Nevertheless, CEPS is rarely applied in academia and industry since it is not broadly applicable and suffers from severe hydrolytic side-reactions. A more commonly used strategy is native chemical ligation[2], which however relies on certain amino acids, e.g. an N-terminal Cys, and is not scalable due to the instability of the peptide thioesters.

CEPS is usually carried out by using (hydrolytic) proteases in the reversed direction[3,4]. Since peptide bond synthesis is thermodynamically unfavourable in the presence of water, the use of neat organic solvents with very low water activity can be beneficial for peptide coupling. For instance, the industrial serine protease Alcalase has been applied for peptide synthesis under nearly anhydrous conditions[5–8]. However, proteases exhibit very low activity in neat organic solvent[8] and the peptide substrates are often poorly soluble both with side-chain protected or

(partially) unprotected. An alternative approach would be performing reactions in water. Here, a peptide C-terminal ester (acyl donor) is coupled to a peptide N-terminal amine (acyl acceptor), and the enzyme or reaction system is designed to kinetically favour the ligation product. There are some effective ligating enzymes known from nature, such as sortase[9] and butelase[10], but they require specific recognition sequence motifs.

Protein engineering has been applied to tailor more broadly applicable proteases for peptide coupling in water. Wells and co-workers constructed subtiligase, a double mutant derived from the *Bacillus amyloliquefaciens* subtilisin BPN'[11,12]. The catalytic serine residue was mutated to a cysteine, and a proline near the active site was converted to an alanine to reduce steric crowding. The hydrolytic activity of this mutant is significantly reduced compared to wild-type, since the introduced cysteine forms a thioester intermediate with the acyl donor, which is more prone to the amine nucleophilic attack leading to the ligation product. Unfortunately, the average ligation yield is 66%[11], even when using at least a ten-fold excess of the acyl acceptor[12], making this technology economically non-viable. Elliott *et al.* reported another approach by using a mutant of *Streptomyces griseus* protease B (SGPB) for peptide coupling[13]. Here, the active site catalytic serine is mutated to an alanine and the reaction proceeds through a proposed histidine-involved acyl-enzyme intermediate. The stability of both enzymes was improved by introducing additional mutations, yielding stabiligase and streptoligase respectively[14,15]. However, in both cases, an enzyme bound cation, is crucial for the stability, which restricts their application. Alternative broadly applicable, robust and cation-free enzymes that catalyze peptide coupling reactions without (hydrolytic) side reactions are highly desired. Since such enzymes have not been discovered, a stable and cation (calcium) independent variant of subtilisin BPN' (Sbt149)[16] was chosen in our study as the starting point for protein engineering. This variant has a deleted calcium binding domain (deletion of 9 amino acids, number 75 to 83) and an additional 18 stabilizing mutations, including one disulphide bridge. We herein report that by introducing the Ser212Cys and the Pro216Ala mutation into this robust enzyme scaffold, a new subtilisin variant is obtained, termed peptiligase, that catalyzes efficient peptide ligation in water.





## Materials and methods

**Chemicals and reagents —** Chemicals and reagents are listed in Chapter 3

**Construction of peptiligase plasmid —** The Sbt149 encoding gene with its natural promoter sequence (a kind gift from prof. P.N. Bryan, University of Maryland, USA) was used as the template to generate the Ser212Cys/Pro216Ala mutant by the megaprimer method. The final PCR product was purified, digested with EcoRI and BamHI, and ligated into the *Eco*RI/*Bam*HI treated pBS-42 vector (DSM8748, DSMZ, Braunschweig, Germany). The ligation mixtures were transformed to competent *E. coli* ER 1821 cells (NEB, Ipswich, MA, USA) and transformants were plated on LB plates containing 34 µg/mL chloramphenicol. The mutant gene was confirmed by DNA sequencing.

**Expression and purification of peptiligase —** Plasmid pBS42-peptiligase was transformed into *B. subtilis* DB104, which is a double mutant (*his nprR2 nprE18 aprΔ3*) strain deficient in extracellular neutral and serine proteases[17] and was a kind gift from prof. O. Kuipers (Molgen). Transformants grown on an LB plate containing 10 µg/mL chloramphenicol at 37°C for 16h were picked and inoculated into 10 mL of LB broth containing 10 µg/mL chloramphenicol. After 16h of incubation at 37°C, 1 % (*v/v*) of the cultures were inoculated to 1 L rich broth (20 g/L yeast extract, 5 g/L trypton, 6.4 g NaCl, 30 mM phosphates pH 7.6, 10 g/L glucose, 50 mg/L Trp, 50 mg/L Lys, 50 mg/L Met, 0.06 mM $MnCl_2$). Cultures were grown at 37°C with shaking and incubation was continued for 48 h. Next, medium was harvested by centrifugation at 6,000 g for 20 min, 4°C. Subsequently, 5 g of $CaCl_2$ were added to the medium to precipitate phosphate salts and the pH was adjusted to 7.5. The precipitate was removed by centrifugation at 6,000 g for 20 min, 4°C. Ammonium sulfate was added to the supernatant to a final concentration of 45wt%. The mixture was stirred for 1 h at 4°C to precipitate peptiligase. Precipitated peptiligase was harvested by centrifugation at 8,000 g for 30 min, 4°C. The pellet was washed 4 times with 100 mL 80% acetone, and resuspended in 15 mL water. Resuspended material was centrifuged at 15,000 g for 10 min, 4°C, to remove insoluble material. The supernatant was desalted using a HiPrep 26/10 desalting column in buffer (20 mM Tricine, pH 7.5). The desalted proteins were loaded on a 5 mL HiTrap Q HP column. Flow through, which contains peptiligase, was collected and concentrated. The purity of the protein was analyzed by SDS-PAGE and estimated to be 70%. The purified enzyme was

flash-frozen in liquid nitrogen and stored at -80°C until further use. Enzyme concentration was determined by measuring absorption at 280 nm.

**Expression and purification of subtiligase —** A gene encoding of subtiligase[8] was provided by DNA2.0 and recloned into EcoRI-BamHI digested *E. coli- B. subtilis* shuttle vector pBS42.The plasmid pBS42-S5 was propagated in *E. coli* strain MM294 (DSM5208, DSMZ, Braunschweig, Germany) isolated and validated by sequencing. The validated plasmid was used for transformation of *B. subtilis* DB104. The production and purification protocol was similar to that for peptiligase, but with the buffers supplemented with 1 mM $CaCl_2$. The yield of purified subtiligase was 2-5 mg/L culture.

**Thermostability assays —** The fluorescence based thermal stability assay[18] was used to determine apparent melting temperatures of the proteins. For this, 20 µl of protein solution in buffer (20 mM Tricine, pH 7.5) was mixed with 5 µl of 100 x Sypro Orange (Molecular Probes, Life Technologies, San Diego, CA) dye in a thin wall 96-well PCR plate. The plate was sealed with Optical-Quality Sealing Tape and heated in CFX 96 Real Time PCR System (BioRad, Hercules, CA, USA) from 20 to 99°C at a heating rate of 1.75°C/min. Fluorescence changes were monitored with a charge-coupled device (CCD) camera. The wavelengths for excitation and emission were 490 and 575 nm, respectively.

**Peptide ligation reactions —** Typical ligation reactions were performed at 25°C in 100 mM Tricine (pH 8.0), containing 15 µM peptiligase and different acyl donors and acyl acceptors. The efficiency of the ligation reactions was analyzed by an LC/MS ion-trapping system (Thermo Scientific) using methods described in the Chapter 5. For reactions using tryptophan containing acyl donor, the UV detection was set at 280 nm, and concentration of the compounds were calculated based on their corresponding peak area. For reactions using tryptophan-free acyl donor, the UV detection was set at 220 nm and the conversion was calculated based on the consumption of the acyl donor substrate. Chemical hydrolysis of the acyl donor was calculated from control experiments.

**pH-ligation profile —** The reaction mixture containing 1 mM Ac-Phe-Ile-Glu-Trp-Leu-OCam, 3 mM Ala-Phe-amide, and 15 µM peptiligase were incubated at 25°C for 1 h. To determine the effect of pH on the ligation efficiency, experiments





were carried out over a pH range of 6.0 – 8.85. The buffers used were MES (100 mM, pH range 6.0-7.0) and Tricine (100 mM, pH range 7.5-8.85). The chemical hydrolysis of the acyl donor was assayed in the absence of the enzyme.

**Effect of organic solvent and additives on peptiligase activity —** The standard reaction mixtures consisted of 1 mM Ac-Phe-Ile-Glu-Trp-Leu-OCam, 3 mM Ala-Phe-amide, 15 µM enzyme and different concentrations of organic solvent, metal ions, chelating agent or denaturing agent in 100 mM Tricine buffer (pH 8.0). The assays were performed at 25°C.

**Synthesis of a 19-mer peptide —** The coupling reaction was performed using 4 mM acyl donor (Ac-Lys-Lys-Lys-Lys-Lys-Lys-Asp-Phe-Ser-Lys-Leu-OCam-Leu-OH) and 6 mM acyl acceptor (H-Ala-Ala-Pro-Arg-Ala-Ala-Arg-Glu-OH) in phosphate buffer (1 M, pH 8.0). To this reaction mixture, 5 µg of enzyme was added and the mixture (1 mL) was analyzed by LC-MS after 30 min. The Cam-ester starting material was fully consumed and the ratio between the product and hydrolysis peak was 97/3. See Supporting Information for further details.

**Cyclization of linear microcin J25 Cam-Leu-ester —** The cyclization reaction was performed with 1 mg linear peptide Cam-ester (H-Gly-Thr-Pro-Ile-Ser-Phe-Tyr-Gly-Gly-Gly-Ala-Gly-His-Val-Pro-Glu-Tyr-Phe-Val-Gly-Ile-OCam-Leu-OH) in 1 mL phosphate buffer (100 mM, pH 8.0) supplemented with DMSO (20 vol%). To this reaction mixture, 10 µg of peptiligase was added and the solution was analyzed by LC-MS after 30 min. Conversion to product was measured by integrating the starting material, product and hydrolyzed Cam-ester peaks. The amount of product after 30 min was 82% and the amount of hydrolysis 18%. The Cam-ester starting material was completely consumed. See Supporting Information for further details. In a second reaction, concentrated substrate solution in DMSO (100 mg/mL) was prepared and dosed in time (20 µl every 15 min) to the enzyme solution (10 µg peptiligase) in 200 µL phosphate buffer (100 mM, pH 8.0). The pH was continuously kept at 8.0 using 5 N aqueous NaOH. The reaction mixture was analyzed by LC-MS after 150 min as described above. The starting material was fully consumed and the amount of product was 81% and hydrolysis 19%.

## Results

The gene encoding peptiligase was constructed by site-directed mutagenesis and cloned into *E. coli/B. subtilis* shuttle vector pBS42. Production of the enzyme was performed in *B. subtilis* DB104 strain which lacks native extracellular neutral and serine proteases expression[12]. Peptiligase is translated as a precursor and needs to be processed to obtain the mature protein. It was reported that a 'helper' protease, i.c. native subtilisin, is required for the processing of the mutant, which has reduced protease activity[8]. In the present study, we observed that peptiligase can perform the autoproteolytic removal of the pro-sequence without any additional subtilisin, which suggests the retained proteolytic activity of this enzyme is sufficient for its maturation. Secreted peptiligase was purified from the culture medium after ammonia precipitation and ion-exchange chromato¬graphy. The yield from 1 liter of cell culture was approximately 20 mg (70% pure on SDS-PAGE gel). The thermostability of the purified peptiligase was determined using the Thermofluor assay. An apparent transition temperature (Tm, app) of 66°C indicates that the enzyme well preserves the thermostability of its parent.

As anticipated, peptiligase is cation-independent. In a standard peptide ligation reaction consisting of activated acyl donor as carboxamidomethyl (Cam) ester (Ac-Phe-Ile-Glu-Trp-Leu-OCam) and amidated acyl acceptor (H-Ala-Phe-$NH_2$) none of the metal cations tested ($Ca^{2+}$, $Mg^{2+}$, $Mn^{2+}$, $Ni^{2+}$) displayed a remarkably stimulatory or inhibitory effect on the activity of peptiligase at 10 mM concentration. The influence of metal ion chelating reagent EDTA (10 mM) on the enzymatic activity was also negligible, in contrast to enzymes without a deleted calcium domain. In addition, the thermostability of peptiligase also remained unaffected (Table 1). The optimal pH value (highest synthetic/hydrolytic ratio) was achieved at pH 8.0. When a higher pH was used, chemical hydrolysis of the peptide ester was more profound. At lower pH the amine acyl donor is protonated and unreactive.

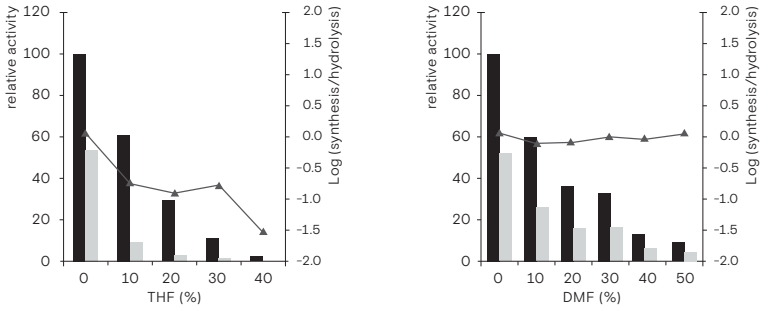


| | Activity (mU/mg) | $T_{m,\,app}$ (°C) |
|---|---|---|
| Control | 26 | 66.0 |
| $Ca^{2+}$ (10 mM) | 20 | 65.5 |
| $Mg^{2+}$(10 mM) | 26 | 65.0 |
| $Mn^{2+}$(10 mM) | 30 | 64.5 |
| $Ni^{2+}$(10 mM) | 22 | 62.0 |
| EDTA(10 mM) | 26 | 66.5 |

**TABLE 1.**

Effect of metal ions and chelating agent EDTA on the (Ac-Phe-Ile-Glu-Trp-Leu-OCam + H-Ala-Phe-$NH_2$) coupling activity and thermostability of peptiligase. Details are given in the Experimental section

To test whether peptiligase tolerates high concentrations of organic solvents, peptide ligation reactions were performed in the presence of various amounts of water-miscible cosolvents, including tetrahydrofuran (THF), *N,N*-dimethylformamide (DMF) and dimethylsulfoxide (DMSO). Reaction mixtures again contained the Ac-Phe-Ile-Glu-Trp-Leu-OCam acyl donor and H-Ala-Phe-$NH_2$ as acceptor. Conversion of substrate and product formation were estimated by HPLC.



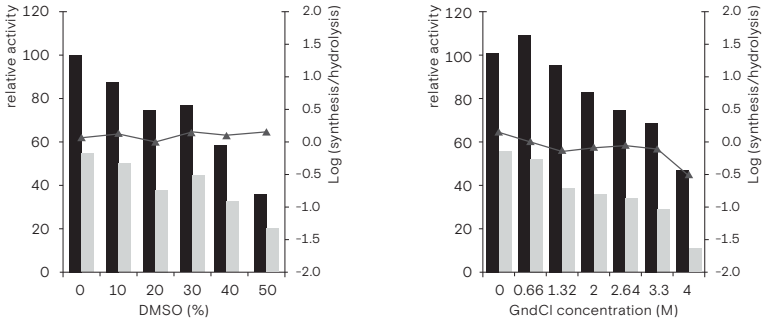



FIGURE 1.

Effect of organic solvents and denaturing agent on the peptiligase catalyzed peptide ligation. Black bars: relative total activity; grey bars: relative synthetic activity; grey triangles: log value of synthesis/hydrolysis ratio.

As shown in Figure 1, the enzyme exhibited moderate tolerance towards THF, and a reasonable residual total activity (of 30%) was observed in the presence of 20 vol% THF. However, THF strongly repressed the nucleophilicity of the acyl acceptor and hydrolysis of the acyl donor became the dominate reaction. Only hydrolysis of the acyl donor Cam ester to free acid was observed. In contrast, the S/H ratio was not significantly affected by the addition of either DMF or DMSO. The enzyme also showed good stability in these two cosolvent systems. Notably, the enzyme retained 36% of its activity in the presence of 50 vol% DMSO in water, which is a very good mixture for dissolving hydrophobic peptides. Therefore, if substrate solubility requires, DMSO should be considered as the preferred organic cosolvent for the coupling reaction.

A possible issue for peptide ligation, as described by Chang *et al.*[14], that substrates could adopt a secondary structure, which could reduce the accessibility to the binding pockets of the coupling enzyme. In such cases, denaturing agents, such as guanidine hydrochloride (GndCl) or urea, can be added to weaken the secondary structure and expose the ligation junction of the substrates. Under extreme denaturing conditions, peptiligase retained significant activity; up to 4 M GndCl could be used. Moreover, the S/H ratio was only marginally influenced by the addition of GndCl. Next, peptiligase was investigated with relation to its peptide ligation efficiency. The parent enzyme Sbt149 binds the substrate in an extended antiparallel β-sheet conformation, while substrate residues flanking the





scissile bond (peptide groups P4-P3') are in close contact with the respective binding pockets (subsites S4-S3')[16]. To test whether ligation reactions catalyzed by peptiligase also require both the acyl donor and the acyl acceptor substantially occupying the binding pockets of the enzyme, we used substrates of varying lengths in peptiligase catalyzed couplings (Tables 2-4). First, single amino acid and dipeptide substrates were coupled via the kinetically controlled conversion. Cam-esters are optimal substrates for proteases/ligases as they mimic a glycine residue[11,12]. The Cam-ester substrates are easily produced via standard solid phase peptide synthesis using conventional resins in high yield and good purity[19]. However, when such amino acids and dipeptides were tested in coupling reactions, the efficiency by peptiligase was poor (Table 2).

| Acyl donor | Acyl acceptor | Conversion (%) | Synthesis (%) | Hydrolysis (%) | S/H ratio |
|---|---|---|---|---|---|
| Cbz-F-OCam | H-A-$NH_2$ | 26 | 3 | 23 | 0.13 |
| Cbz-F-OCam | H-AF-$NH_2$ | 50 | 6 | 44 | 0.13 |
| Cbz-AF-OCam | H-A-$NH_2$ | 53 | 2 | 51 | 0.04 |
| Cbz-AF-OCam | H-AF-$NH_2$ | 53 | 6 | 47 | 0.13 |

TABLE 2.

Peptiligase catalyzed coupling reactions with short peptide substrates. Reactions were performed with 1 mM acyl donor and 3 mM acyl acceptor in 100 µl Tricine buffer (100 mM, pH 8.0). Reaction time =3 h

Subsequently, a series of longer (5-11 amino acids) acyl donor Cam-esters with the same ligation junctions were tested. The results (Table 3) indicate that peptiligase catalyzes the coupling reaction of longer acyl donors with high efficiency, and that there is a clear correlation between the length of the acyl donor and the S/H ratio of the reaction. Virtually no hydrolytic side-reactions were observed at all when an 11-mer Cam-ester was used, i.e. an S/H ratio of more than 100. This corresponds to a conversion of >99% to product and less than 1% of hydrolysis. Notably, this high yield was obtained using only 1.5 equivalent of acyl acceptor, making this an economically very attractive ligation technology.

| Acyl donor | Conversion (%) | Synthesis (%) | Hydrolysis (%) | S/H ratio |
|---|---|---|---|---|
| Ac-DLSKQ-OCam | 93 | 81 | 12 | 6 |
| Ac-TSDLSKQ-OCam | 85 | 83 | 2 | 41.5 |
| Ac-TFTSDLSKQ-OCam | 66 | 64 | 2 | 32 |
| Ac-EGTFTSDLSKQ-OCam | 60 | 60 | <1 | >100 |

**TABLE 3.**

Peptiligase-catalyzed coupling reactions using acyl donors with varying lengths and H-AF-$NH_2$ as the amine nucleophile. Reactions were performed with 10 mM acyl donor and 15 mM acyl acceptor in 100 µl Tricine buffer (100 mM, pH 8.0). Reaction time =3 h.

Next, we investigated whether the length of the acyl acceptor has a similar effect (Table 4). The efficiency of peptide ligation was analyzed using a series of acyl acceptors, with the same amino-terminal residue and subsequent length extension (1–4 amino acids). While a coupling reaction with alanine amide as nucleophile appeared to be accompanied by overwhelming hydrolysis, longer peptide amides served as better acyl acceptors.

| Acyl acceptor | Conversion (%) | Synthesis (%) | Hydrolysis (%) | S/H ratio |
|---|---|---|---|---|
| H-A-$NH_2$ | 35 | <1 | 35 | <0.01 |
| H-AF-$NH_2$ | 59 | 32 | 27 | 1.2 |
| H-AFA-$NH_2$ | 25 | 13 | 11 | 1.2 |
| H-AFAY-$NH_2$ | 23 | 14 | 9 | 1.5 |

**TABLE 4.**

Peptiligase-catalyzed coupling reactions using amine nucleophiles with varying length and Ac-FIEWL-OCam as acyl donor. Reactions were performed with 1 mM acyl donor and 3 mM acyl acceptor in Tricine buffer (100 mM, pH 8.0). Reaction time =1 h.





These results suggest that occupation of both the S1' and the S2' subsites as well as the S1-S4 subsites is a prerequisite for efficient ligation. Thus, the minimal peptide length for an efficient coupling reaction is 4 + 2, with longer segments giving more efficient ligation, as desired. From the results of an 11 + 8 coupling reaction, i.e. Ac-Lys-Lys-Lys-Lys-Lys-Lys-Asp-Phe-Ser-Lys-Leu-OCam-Leu-OH + H-Ala-Ala-Pro-Arg-Ala-Ala-Arg-Glu-OH, it became clear that by far we have not reached the limitations of this technology since the reaction reached 97% conversion to 19-mer product with only 3% of Cam-ester hydrolysis (details in supporting information). Clearly, this enzymatic ligation technology can be used for the synthesis of medium-sized and long peptides.

To determine whether the synthetic properties of peptiligase are better than those of subtiligase, the S/H ratio was determined using various acyl acceptors. Sub-optimal (diluted) reaction conditions were used to clearly quantify the amounts of synthetic product and product of the hydrolytic side reaction.

Under the same reaction conditions, the S/H ratio of peptiligase is surprisingly higher than that of subtiligase; in most cases the S/H ratio of peptiligase was twofold higher. The improved S/H ratio together with the excellent stability render peptiligase as a highly promising enzyme for further development of a generally applicable CEPS technology. We are currently investigating the peptiligase-catalyzed synthesis of various pharmaceutical products on gram scale to determine purified product yield and compare the results to conventional synthetic methods.

| | Peptiligase | | Subtiligase | |
|---|---|---|---|---|
| Acyl acceptor H-Xxx-Leu-Arg-$NH_2$ | Synthesis (%) | S/H ratio | Synthesis (%) | S/H ratio |
| Ser | 58 | 1.38 | 40 | 0.67 |
| Gly | 48 | 0.92 | 25 | 0.33 |
| Ala | 47 | 0.89 | 20 | 0.25 |
| Asp | 21 | 0.27 | 6.8 | 0.07 |
| Acyl acceptor H-Ala-Xxx-Arg-$NH_2$ | Synthesis (%) | S/H ratio | Synthesis (%) | S/H ratio |
| Ile | 46 | 0.83 | 26 | 0.35 |
| Val | 46 | 0.86 | 29 | 0.41 |
| Leu | 32 | 0.46 | 26 | 0.34 |
| Met | 31 | 0.38 | 24 | 0.28 |

[a] Reactions were performed with 0.75 mM acyl donor and 1.5 mM acyl acceptor in Tricine buffer (100 mM, pH 8.0) and 5 µg of enzyme. Full conversion was reached after 30 min. [b] Products yields derived from HPLC analysis of reaction mixtures.

**TABLE 5.**

P1'and P2' substrate scope of peptiligase and subtiligase using Ac-DFSKL-OCam as acyl donor and various acyl acceptors, where Xxx = the proteinogenic amino acid indicated. Best eight results shown.

Besides peptide segmentt condensation, subtiligase has been used to cyclize linear peptides longer than 13 amino acids in reasonable yield (average of 64%)[20]. Macrocylic peptides are a very promising class of pharma¬ceuticals, mainly due to their high stability in blood serum[20]. We embarked on the cyclization (Figure 2) of linear Microcin J25, a 21 amino acid long lasso-peptide naturally produced by a strain of *E. coli* isolated from human feces with antibiotic properties[21]. Chemical head-to-tail cyclization of such a long peptide is very challenging and the reaction has to be performed with extremely diluted substrate to prevent polymerization, making it difficult for scale-up[22]. Microcin J25 is hydrophobic and cosolvents such as DMSO are needed in high amounts to solubilize the peptide in aqueous environment, which is well tolerated by peptiligase.





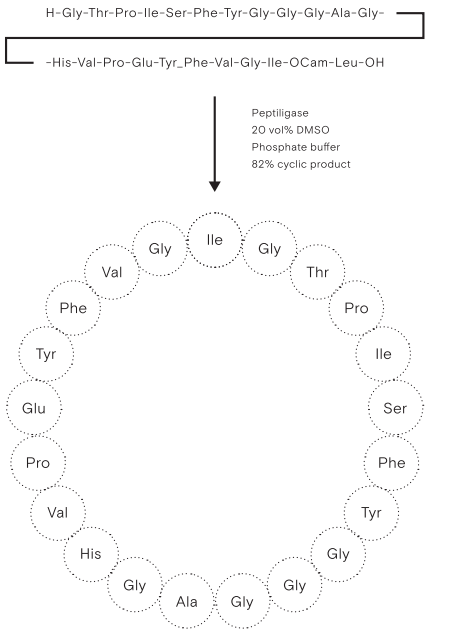



**FIGURE 2.**

Head-to-tail cyclization of linear Microcin J25 (21-mer) by peptiligase in 20% (*v/v*) DMSO (details in Supporting Information).

When activated linear Microcin J25 (1 mg/mL) was cyclized with peptiligase using 20% (v/v) DMSO in phosphate buffer, a very good conversion to cyclic product was observed, i.e. 82% of synthetic product and only 18% of hydrolysis of the linear Microcin J25Cam ester to free acid. No dimerization or polymerization was detected. Moreover, a similar conversion to cyclic peptide (81%) was obtained in a reaction with concentrated substrate (100 mg/mL in DMSO) dosed to peptiligase in phosphate buffer, thus obtaining a concentrated product solution (50 mg/mL). Clearly, besides peptide segment condensation, peptiligase can be used to synthesize head-to-tail macrocyclic peptides in an efficient manner.

## Conclusion

We have successfully constructed a stable peptide ligase that is able to ligate peptides in water with very high S/H ratios. When longer peptides are coupled, the hydrolytic side-reaction is almost completely suppressed. To the best of our knowledge, this is the first reported broadly applicable and traceless peptide ligating enzyme that is independent of cations and is very robust. The enzyme can be expressed in *B. subtilis* and can be directly isolated from the culture medium since it is secreted and no additional protease is required for maturation. Furthermore, peptiligase is compatible with a variety of additives and cosolvents, has a relaxed specificity in coupling reactions, and does not hydrolyze amide bonds in the main chain of precursor or product peptides. These features endow the enzyme with catalytic properties required for the synthesis of a variety of medium-sized and long (pharmaceutical) peptides. Moreover, we have shown that peptiligase can be used for the synthesis of a peptide macrocycle, which represents an important upcoming class of pharmaceutical peptides for future medicines. Based on the robust framework of peptiligase, further protein engineering for even higher synthetic performance or broader substrate acceptance is ongoing.

**Acknowledgments**

This project is part of Integration of Biosynthesis and Organic Synthesis program (IBOS-2; program number: 053.63.014) funded by The Netherlands Organisation for Scientific Research (NWO) and Advanced Chemical Technologies for Sustainability (ACTS). The authors thank Dr. J. M. van der Laan from DSM for helpful discussions.

**Author contribution**

AT, BW and TN did experimental work. AT, TN, BW and DBJ wrote the paper. PJLMQ and DBJ supervised the work.





## References

01 — Guzman, F., Barberis, S. & Illanes, A. Peptide synthesis: chemical or enzymatic. *Electron. J. Biotechnol.* **10**, 279–314 (2007).

02 — Dawson, P. E., Muir, T. W., Clark-Lewis, I. & Kent, S. B. Synthesis of proteins by native chemical ligation. *Science* **266**, 776–779 (1994).

03 — Bongers, J. & Heimer, E. P. Recent applications of enzymatic peptide synthesis. *Peptides* **15**, 183–93 (1994).

04 — Bordusa, F. Proteases in organic synthesis. *Chem. Rev.* **102**, 4817–68 (2002).

05 — Nuijens, T., Schepers, A. H. M., Cusan, C., Kruijtzer, J. A. W., Rijkers, D. T. S., Liskamp, R. M. J. & Quaedflieg, P. J. L. M. Enzymatic fragment condensation of side chain-protected peptides using Subtilisin A in anhydrous organic solvents: A general strategy for industrial peptide synthesis. *Adv. Synth. Catal.* **355**, 287–293 (2013).

06 — Nuijens, T., Cusan, C., Schepers, A. H. M., Kruijtzer, J. A. W., Rijkers, D. T. S., Liskamp, R. M. J. & Quaedflieg, P. J. L. M. Enzymatic synthesis of activated esters and their subsequent use in enzyme-based peptide synthesis. *J. Mol. Catal. B Enzym.* **71**, 79–84 (2011).

07 — Vossenberg, P., Beeftink, H. H., Nuijens, T., Cohen Stuart, M. A. & Tramper, J. Selecting optimal conditions for Alcalase CLEA-OM for synthesis of dipeptides in organic media. *J. Mol. Catal. B Enzym.* **75**, 43–49 (2012).

08 — Klibanov, A. M. Why are enzymes less active in organic solvents than in water? *Trends Biotechnol.* **15**, 97–101 (1997).

09 — Ritzefeld, M. Sortagging: a robust and efficient chemoenzymatic ligation strategy. *Chemistry* **20**, 8516–29 (2014).

10 — Nguyen, G. K. T., Wang, S., Qiu, Y., Hemu, X., Lian, Y. & Tam, J. P. Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis. *Nat. Chem. Biol.* **10**, 732–8 (2014).

11 — Braisted, A., Judice, J. & Wells, J. A. Synthesis of proteins by subtiligase. *Methods Enzymol.* **289**, 298–313 (1997).

12 — Abrahmsén, L., Tom, J., Burnier, J., Butcher, K. A., Kssiakoff, A. & Wells, J. A. Engineering subtilisin and its substrates for efficient ligation of peptide bonds in aqueous solution. *Biochemistry* **30**, 4151–9 (1991).

13 — Elliott, R. J., Bennet, A. J., Braun, C. a, MacLeod, A. M. & Borgford, T. J. Active-site variants of Streptomyces griseus protease B with peptide-ligation activity. *Chem. Biol.* **7**, 163–171 (2000).

14 — Chang, T. K., Jackson, D. Y., Burniert, J. P. & Wells, J. A. Subtiligase : A tool for semisynthesis of proteins. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 12544–12548 (1994).

15 — Joe, K., Borgford, T. J. & Bennet, A. J. Generation of a thermostable and denaturant-resistant peptide ligase. *Biochemistry* **43**, 7672–7 (2004).

16 — Ruan, B., London, V., Fisher, K. E., Gallagher, D. T. & Bryan, P. N. Engineering substrate preference in subtilisin : Structural and kinetic analysis of a specificity mutant. *Biochemistry* **47**, 6628–6636 (2008).

17 — Kawamura, F. & Doi, R. H. Construction of a *Bacillus subtilis* double mutant deficient in extracellular alkaline and neutral proteases. *J. Bacteriol.* **160**, 442–444 (1984).

18 — Ericsson, U. B., Hallberg, B. M., Detitta, G. T., Dekker, N. & Nordlund, P. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal. Biochem.* **357**, 289–98 (2006).

19 — de Beer, R. J. A. C., Nuijens, T., Wiermans, L., Quaedflieg, P. J. L. M. & Rutjes, F. P. J. T. Improving the carboxyamidomethyl ester for subtilisin A-catalysed peptide synthesis. *Org. Biomol. Chem.* **10**, 6767–75 (2012).

20 — Jackson, D. Y., Bumier, J. P. & Wellst, J. A. Enzymatic cyclization of linear peptide esters using subtiligase. *J. Am. Chem. Soc.* **117**, 819–820 (1995).

21 — Wilson, K. A., Kalkum, M., Ottesen, J., Yuzenkova, J., Chait, B. T., Landick, R., Muir, T., Severinov, K. & Darst, S. A. Structure of microcin J25, a peptide inhibitor of bacterial RNA polymerase, is a lassoed tail. *J. Am. Chem. Soc.* **125**, 12475–12483 (2003).

CHAPTER — 08

# Summary and outlook

Ana Toplak and Dick B. Janssen

The focus of this thesis is discovery and engineering of proteases for peptide synthesis. Using genome mining and heterologous expression, novel thermostable subtilisin-like proteases originating from extremophiles were produced and their properties are described, including their activity in peptide synthesis. The novel biocatalysts could be applied under harsh reaction conditions, i.e. at high temperature, in the presence of organic cosolvents or in anhydrous organic solvents, and in the presence of denaturants. In addition, engineering of a stable, cation-independent thiol-subtilisin for efficient peptide synthesis in water is reported. Finally, we discuss the future perspectives of chemoenzymatic peptide synthesis technology.

## Summary

A generic technology for the production of peptides does not exist. Current methods for peptide production are chemical synthesis, production in recombinant systems, and chemo-enzymatic synthesis (CEPS), and the preferred synthetic route depends on sequence, length and properties of the target peptide[1]. A short description of the methods is given in Chapter 1. At this moment, chemical peptide synthesis is still the preferred method for the production of therapeutic peptides, since it allows the synthesis of almost any desired peptide sequence, including peptides with non-proteinogenic amino acids. Usually, in the early peptide drug development phase the synthesis is performed on a solid-phase, where a large excess of reagents is required to reach high yields and large amounts of solvent are used in washing steps. As the complexity and length of the target peptide grow, tedious and expensive preparative HPLC purification is needed to remove impurities, which further increases production costs. In addition, longer peptides (>10 amino acids) tend to form tertiary structures due to a process called hydrophobic collapse, making further synthetic steps difficult. Therefore, a hybrid approach where peptide segments are synthesized on solid-phase and chemically coupled in solution is the usual way to produce longer peptides. One needs to carefully select ligation positions, i.e. the segments obtained by solid-phase methods should be of similar size and have Gly or Pro at the C-terminus to avoid racemization. Side chains should be adequately protected, and the peptides should be soluble. If racemization occurs or side chains react, product purification becomes troublesome and expensive. An elegant chemical method for synthesis of long peptides (>200 amino acids) is native chemical ligation[2]. It allows coupling of unprotected peptide segments in buffered solutions.





However, it is not applicable on industrial scale due to the instability of thioester precursors and long reaction times with diluted substrate solutions. Due to the low environmental impact, ribosomal expression and production of the recombinant peptides by fermentation would be a promising low-cost green technology for large-scale peptide production. However, high development costs are involved because production strains need to be developed and isolation must be optimized for each new peptide sequence. Also, there is a sequence limitation because only proteinogenic amino acids can be incorporated in a straightforward manner[3]. As a result, ribosomal peptide synthesis is only explored in the late drug development stage. Another recombinant method sometimes applied on industrial scale involves non-ribosomal peptide synthetases (NRPs). These multimeric enzyme complexes are naturally involved in the production of a large diversity of short peptides, which very often contain non-proteinogenic amino acids. These systems are exploited for production of antibiotics such as β-lactams and cyclosporins[4]. Although modification of the peptide scaffold by engineering NRPS modules or domains is possible, such as incorporation of D-amino acids and peptide cyclization, the engineering of an NRPS for production of a specific target peptide remains a complicated task. Chemoenzymatic peptide synthesis (CEPS) is an attractive alternative approach for the production of medium to long-sized peptides. Enzymatic coupling of chemically synthesized peptide segments combines advantages of solid-phase synthesis (e.g. incorporation of non-natural amino acids) with enzymatic ligation of peptide segments (lack of racemization, no need for protection groups, use of an aqueous phase with better solubility of peptides). Enzymatic ligation reactions in CEPS are also regio- and stereospecific, thereby reducing the necessity of expensive and troublesome HPLC purification steps and reducing overall production costs. Development of improved peptide coupling enzymes is crucial for the future success of chemoenzymatic peptide synthesis as addressed in this thesis. Indeed, since the properties of the catalyst determine to a large extend the feasibility of practical application, the discovery and engineering of better proteases is an intensive field of research. In this project we have chosen a genome mining approach to obtain novel enzymes for peptide synthesis. Chapter 2 gives an overview of the properties of known enzymes that have been used in peptide synthesis and activation. In general, hydrolytic enzymes such as proteases can be employed as catalysts for peptide bond formation either under thermodynamic or kinetic control[5]. Thermodynamic control is a reversal of hydrolysis and all proteases regardless of their mechanistic class can ligate peptides according to this principle. An example is the thermolysin –catalyzed production of aspartame on industrial scale[6]. As an alternative, proteases of the serine or cysteine hydrolase class can form a covalent bond with an activated precursor (acyl donor) and

catalyze peptide synthesis following kinetic principles. This requires that the acyl-enzyme intermediate is preferentially cleaved by a nucleophilic peptide that acts as acyl acceptor. Reactions proceeding under kinetic control tend to be faster and higher conversions can be achieved with less enzyme than during thermodynamically controlled conversions. A variety of alternative nucleophiles such as amides, peroxides, hydrazine, alcohols or hydroxamic acids can be used as acyl acceptor, thereby expanding the biocatalytic scope of these proteases[7]. Many classical proteases have been successfully used in peptide synthesis, but intensive medium and substrate engineering are required. The latter involves selection of coupling sites for peptide segment condensation, and selection of leaving groups if kinetically controlled synthesis is applied. Strongly activating leaving groups, i.e. carboxamidomethyl (OCam) and its derivatives, can expand the application scope of proteases[8]. Some reactions can be done in the presence of cosolvents to increase peptide solubility or in neat organic solvent to suppress peptide hydrolysis. Neat organic solvents may change enzyme selectivity, and broaden the range of reactions that can be performed[9]. However, enzyme activity in neat organic solvents is usually very low[10] and efficient conversion requires solvent-compatible enzymes which show both decent stability and activity under near-anhydrous conditions. The main side reaction of protease-catalyzed peptide synthesis under aqueous conditions is hydrolysis. The goal of the reaction and enzyme engineering studies described in other chapters is to maximize the versatility of coupling reactions while keeping hydrolytic side reactions minimal.

In Chapter 3 we explored the use of genome mining for discovery of thermostable homologs of the classical protease subtilisin E. Subtilisins were selected because of their broad specificity and expected stability. It was assumed that thermophilic organisms contain enzymes that can tolerate harsh conditions such as elevated temperature or exposure to cosolvents and that may have increased activity in neat organic solvents as compared to their mesophilic counterparts. In particular hyperthermophiles, thermophiles and some bioremediation organisms were selected as target organisms. From 73 hits (36 organisms) of genes encoding putative proteases, 50 were selected as possible targets as the properties of the encoded enzymes were not previously described. Cloning and expression were performed in a mesophilic host that is easy to grow (*Escherichia coli*) and for which suitable cloning techniques exist. Another advantage of using a mesophilic host is that enzyme isolation can be based on the difference in thermostability between host proteins and the target thermozyme: a single heat treatment step may be used for purification (Chapter 4). Although successful in the sense that we discovered new enzymes, it was also evident that individual expression and testing of putative proteases identified by genome mining required a considerable experimental effort, with rather unpredictable revenues. Better bioinformatic





tools could possibly focus experimental work on the most promising enzymes. Unfortunately, at this moment it is still impossible to predict directly from the protein sequence or via homology modeling whether a putative subtilase will be expressed and active in a selected host, and if the catalytic properties of such an enzyme will be attractive. In our expression experiments two bottlenecks occurred. First, cloning of GC-rich sequences by PCR amplification and restriction cloning was successful for only 46% of the genes. With the progress of affordable DNA synthesis this problem will be less important in the future. Protein processing and secretion were other critical steps, and these will remain important for obtaining active protease. A plate screening protocol for functional expression in *E.coli* was applied; it is based on including a protease substrate into plates (skim-milk agar plates) on which colonies producing active subtilase could be recognized by formation of a clearance halo. In addition, small-scale liquid culture screening was developed to identify production conditions for novel subtilases. In each screening method, a heating step was applied to enable protease activation by processing. The full preprosequence of 24 putative subtilases was cloned into pBAD or pET-based vectors. From these 24 cloned genes in *E. coli*, nine active subtilases were detected by plate screening and one enzyme was found in small-scale liquid culture screening. Although they were expected to be promising as hyperthermophilic proteases, none of the archaeal subtilases could be obtained by recombinant expression in *E. coli* cells. If produced in *E. coli*, subtilase expression was modest in all cases (5-50 mg per L of culture). Despite these limitations, in total ten putative extremophilic subtilases were expressed using *E. coli* and found to be active. In the following chapters, the expression and biochemical properties of four novel subtilases are described.

In Chapter 4 biochemical characterization of a novel thermostable and cosolvent-tolerant serine protease from *Coprothermobacter proteolyticus* termed proteolysin is reported. Recombinantly produced in *E. coli*, proteolysin could be purified to near homogeneity from crude cell lysate by a single heat-treatment step. Higher levels (20 mg/L) where obtained with maltose-binding protein tag (MBP-tag) fused to the N-terminus of the target enzyme. Although several enzyme forms with different degrees of processing could be identified in the total cell lysate, the heating step led to a single protein band. During heat-processing the C-terminal hexahistidine tag and part of the linker were autocatalytically removed, leading to the conclusion that the C-terminal tag cannot be used to facilitate the purification. The enzyme has a broad pH tolerance and is active at temperatures of up to 80°C, much higher than optimal growth temperature of *C. proteolyticus* (63°C). In addition, proteolysin shows good activity and stability in the presence of organic cosolvents, detergents and dithiothreitol, and remains active in 6 M

guanidinium hydrochloride. The findings suggest a positive correlation between thermostability and cosolvent tolerance. The function of the putative Cys182-Cys201 disulfide bond was investigated. By disrupting the disulfide bond by replacing cysteines with alanine, a more active variant with similar thermostability was obtained, albeit in lower yields. The enzyme prefers hydrophobic amino acid residues at the P1 substrate-binding subsite, and is a candidate for applications in proteomics where protein digestion under harsh conditions is required. The efficiency of this protease in peptide synthesis in neat organic solvent was regrettably not sufficient for further exploration.

Chapter 5 addresses chemoenzymatic peptide synthesis in neat organic solvents. Two novel thermostable subtilisin-like proteases obtained by genome mining, *Dg*Sbt from *Deinococcus geothermalis* and *Taq*Sbt from *Thermococcus aquaticus* Y51MC23. Both were recombinantly expressed in *E. coli* at 25-50 mg/L culture and characterized. The enzymes were highly thermostable (apparent unfolding temperatures 90°C and 70°C, respectively) and a single heat-treatment step yielded purified enzyme. The catalytic efficiencies in standard hydrolytic reactions were one order of magnitude lower than of subtilisin Carlsberg and similar to values observed for other thermostable subtilases. *Taq*Sbt and *Dg*Sbt also showed exceptional stability in the presence of 20 mM EDTA, 10% (*w/v*) SDS, Tween 20 (10% *w/v*), urea and guanidinium hydrochloride (GdmHCl, s6M). The tolerance of these new subtilases to chaotrophic agents was superior to the recently engineered chaophilic subtilisin E variant for the use in diagnostic kits[11]. Hyperthermophilic Tk-SP from the archaeon *Thermococcus kodakaraensis* is another reported protease with high thermostability, but low expression levels (0.8 mg/ L culture[12]) limit further biocatalytic application of Tk-SP. In addition, the novel subtilases *Dg*Sbt and *Taq*Sbt catalyzed efficient peptide bond synthesis at 80°C and 60°C in neat acetonitrile with excellent conversion (>90%). Active enzyme preparations for use in organic solvents were prepared directly from cleared heat-treated cell lysates, thus minimizing the downstream processing by precipitation and rinsing with isopropanol, followed by drying. The enzymes also tolerated high levels of *N,N*-dimethylformamide (DMF) as a cosolvent (40%-50% *v/v*), which improved substrate solubility and gave good conversion in 5+3 peptide condensation reactions. In conclusion, *Dg*Sbt and *Taq*Sbt are useful biocatalysts for short peptide synthesis when high temperature and high DMF concentrations are required. For industrial application of these subtilases in peptide synthesis, their expression level has to be further improved, since rather large amounts of IPRED are required (20% (*w/v*) and higher) for good conversion. A promising start may be production in *Pichia pastoris*, where an expression level of 1g/L was achieved for another thermophilic subtilase called aqualysin I[13].



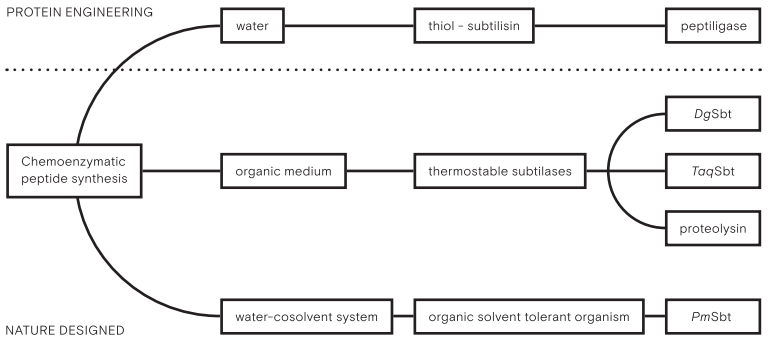


Enzymes with high tolerance towards organic solvents could expand the applicability of biocatalysts in various industrial processes. In Chapter 6 we report recombinant production in *E. coli*, purification and biochemical characterization of the serine protease from *Pseudomonas mendocina* strain ymp, termed *Pm*Sbt. The precursor of *Pm*Sbt has 38% sequence identity to subtilisin E and consists of an N-terminal presequence and prodomain, the mature domain and a C-terminal domain. With the incorporated C-terminal hexahistidine tag, purification using immobilized metal affinity chromatography was possible with a yield of 10 mg of purified protein/L of medium. Deletion of the C-terminal domain (*Pm*Sbtdel) revealed its role in facilitating enzyme secretion in the *E. coli* host. This calcium-activated mesophilic enzyme had an optimal pH of 8.0 and a temperature optimum at 46°C. Out of three putative disulfide bonds, only the S-S bond in the C-terminal domain (Cys479-Cys490) could be disrupted without a negative effect on functional expression. Both *Pm*Sbt and *Pm*Sbtdel aggregate unless glycerol is added to 20% (*v/v*). Bioinformatic analysis and structural inspection identified two segments in the mature domain that may be responsible for the aggregation propensity. The isolated protease *Pm*Sbt exhibits remarkable stability in DMSO and DMF (50% *v/v*) for as long as 14 days. For *Pm*Sbt, the extraordinary stability in organic solvents was not accompanied by a high thermostability as found for other thermostable proteases described in this thesis.

We tested the synthetic properties of *Pm*Sbt on the hexapeptide precursors of the bioactive PHM27 peptide, an agonist of the human calcitonin receptor involved in calcium homeostasis[12]. Surprisingly, *Pm*Sbt showed kinetic preference for water as nucleophile in the presence of 50% (*v/v*) of DMF or DMSO, leading to fast ester hydrolysis, after which a thermodynamically controlled peptide coupling occurred. Thus, in the presence of organic cosolvent (50% *v/v* DMF or DMSO) the equilibrium was shifted to the synthetic product with 60% conversion after 48 h. Due to the organic solvent tolerance, *Pm*Sbt can accept non-activated precursors in thermodynamically controlled peptide synthesis reactions in higher yields than reported for other serine proteases (i.e. trypsin)[13]. With required prolonged reaction times the chance for byproduct formation grows. Besides medium engineering to improve the peptide product yield, also engineering studies aimed in understanding and improving catalytic properties of the *Pm*Sbt would be desired. From the bioinformatics studies segments with high aggregation tendency and three loops forming the substrate-binding site were suggested as targets for the further engineering studies. When synthesizing peptides in aqueous conditions, the use of thiol-subtilisins is advantageous and allows efficient segment coupling. Chemoenzymatic peptide synthesis in aqueous

medium has many advantages: the solubility of unprotected peptides is excellent and there is no need for side-chain protective groups. Chapter 7 is dedicated to peptiligase, a highly engineered subtilisin BPN' that is capable of peptide synthesis in aqueous media. The production, characterization, and the substrate scope are reported. The enzyme is a variant of a hyperstable calcium-independent mutant of subtilisin BPN', with the catalytic Ser212 mutated to Cys and Pro216 converted to Ala. Peptiligase catalyzes exceptionally efficient peptide coupling in water with a surprisingly high synthesis over hydrolysis (S/H) ratio. The S/H ratio of the peptide ligation reaction is correlated to the substrate length and proved to be >100 for the synthesis of a 13-mer peptide, which corresponds to >99% conversion to the ligated peptide product and <1% hydrolytic side-reactions. Furthermore, peptiligase does not require a particular recognition motif resulting in a broadly applicable and traceless peptide ligation technology. Peptiligase is very robust, easy to produce in *Bacillus subtilis*, and its purification is straightforward. It shows good activity and stability in the presence of organic cosolvents and chelating or denaturing agents, enabling the ligation of poorly soluble (hydrophobic) or folded peptides. In addition, peptiligase is able to efficiently catalyze head-to-tail peptide cyclization reactions, as shown for Microcin J25 cyclic variant. This enzyme could be useful for the industrial synthesis of diverse (pharmaceutical) peptides.

To summarize, five novel subtilases are characterized in this thesis, and four of them could be applied in CEPS under different conditions. A biocatalyst for each reaction medium (water, neat organic solvent and water-cosolvent system) is provided (Figure 1).








# Outlook

In the past decade the field of chemoenzymatic peptide synthesis has been oriented towards the use of hydrolases in nonaqueous environments, focusing on minimizing the main side reaction, (i.e. hydrolysis of substrates or product). Approaches using substrate or medium engineering have been in demand for novel organic-solvent tolerant subtilases. Several novel subtilases have been selected in this work and were tested in hydrolytic and synthetic reactions under a variety of conditions. Thermostable subtilases obtained by genome mining of the extremophiles proved to be robust enzymes that can be produced in a mesophilic host and can be easily purified in a single heat-processing step. As such, they can be used as hydrolases under harsh reaction conditions or they can be precipitated by an organic solvent and used as powder preparations for the synthesis of short peptides in mixtures of anhydrous organic solvents (Chapters 4-5). Of particular interest for various applications is the novel subtilase described in Chapter 6 with superb tolerance to DMF and DMSO, allowing thermodynamically controlled peptide synthesis. Despite recent advances in the nonaqueous enzymology, the possibility of enzymatic peptide coupling in aqueous solution would offer many advantages. Increased peptide solubility in buffered solutions and no side-chain protective groups would allow enzymatic coupling of very long peptide segments. Thiol-subtilisins[16] and selenol-subtilisins[17] have been designed as the ligases for the aqueous chemo-enzymatic peptide synthesis.

From the new biocatalysts presented in this thesis, peptiligase, a highly-engineered thiol-subtilisin has the potential to establish chemoenzymatic peptide synthesis technology as the preferred method for synthesis of (pharmaceutical) peptides and proteins in a cost-efficient and environmental friendly method. We have demonstrated high coupling efficiencies (high S/H ratios) of peptiligase-catalyzed ligations, especially for long unprotected peptides in buffered solutions. These kinetically controlled reactions employing an activated acyl donor and a suitable nucleophilic acyl acceptor are fast and require only a small amounts of enzyme. The acyl donor is usually protected at the N-terminus and activated

**FIGURE 1.**

Biocatalysts for chemoenzymatic peptide synthesis explored in this thesis.

at its C-terminus as Cam-Leu ester[8]. The nucleophile used is in the form of an amide. Peptiligase has six substrate binding pockets that play a role in substrate recognition and catalysis. The next step would be to investigate the substrate range of peptiligase and to improve the scope accordingly by enzyme engineering methods such as structure-guided rational design, directed evolution and computational design. It is highly unlikely that one single peptiligase variant will efficiently couple all possible combinations of the substrate amino acid sequence space, thus engineering will focus on producing a toolbox of ligases covering the diversity of practically relevant ligation sequences.

One could describe those enzymes as "restriction ligases", designed for a specific combination of acyl-donor and nucleophile. Such enzymes could allow one-pot peptide assembly reactions and beside applications in pharmaceutical industry, they might be used in synthesis of novel materials, in particular self-assembling biopolymers[18].

Besides synthesis of the known peptide sequences, modification reactions leading to diversification in structure and biological behavior are also of increasing interest. Modified synthetic peptides could act as prodrugs or as dual-action pharmaceuticals. Bioactivity and pharmacokinetics may be modified by making peptide conjugates, peptides containing either D-amino acids or β-amino acids, or by designing cyclic and constrained peptides[19]. As shown in Chapter 7, peptiligase is able to cyclize an 21-mer linear peptide exclusively as a head-to-tail cyclic form using high substrate concentrations and short reaction times. Other attractive applications would emerge if peptiligase could be redesigned for various recognition sequences allowing traceless peptide sequence modifications and labeling. Such traceless peptide modifications are currently impossible using ligases like sortase[20], butelase[21] or trypsiligase[22] since they require specific recognition sequence motif. Selective ligases will also make it possible to couple unprotected peptide segments produced by fermentation, opening the door towards protein- and antibody-drug conjugates.

In conclusion, protein engineering of subtilisin and thiol-subtilisin variants has considerable protential, even though almost 25 years have passed since subtiligase was discovered. The peptide ligase (peptiligase) described in this thesis is a robust, thermostable enzyme with high synthesis/hydrolysis ratio, potentially providing a key to successful industrial chemoenzymatic peptide synthesis technology. Further engineering studies combined with the ever improving possibilities of *in-silico* design should enable tailoring of the biocatalysts to fit the reaction needs.





## References


01 — Guzman, F., Barberis, S. & Illanes, A. Peptide synthesis: chemical or enzymatic. *Electron. J. Biotechnol.* **10**, 279–314 (2007).

02 — Dawson, P. E., Muir, T. W., Clark-Lewis, I. & Kent, S. B. Synthesis of proteins by native chemical ligation. *Science* **266**, 776–779 (1994).

03 — Lax, R. The future of peptide development in the pharmaceutical industry. *PharManufacturing Int. Pept. Rev.* 10–15 (2010).

04 — Felnagle, E. A., Jackson, E. E., Chan, Y. A., Podevels, A. M., Berti, D., Mcmahon, M. D. & Thomas, M. G. Nonribosomal peptide synthetases involved in the production of medically relevant natural products. *Mol Pharm* **5**, 191–211 (2008).

05 — Bordusa, F. Proteases in organic synthesis. *Chem. Rev.* **102**, 4817–4868 (2002).

06 — Quaedflieg, P. J. L. M., Sonke, T. & Wagner, A. F. V. Synthesis and recovery of aspartame involving enzymatic deformylation step.US patent 6,617,127 B2 filed 22 Jun. 2001, and issued 9 Sept. 2003.

07 — Faber, K. *Biotransformations in Organic Chemistry. Ed* **5**. (Springer-Verlag, Berlin, Germany).

08 — de Beer, R. J. A.C., Nuijens, T., Wiermans, L., Quaedflieg, P. J. L. M. & Rutjes, F. P. J. T. Improving the carboxyamidomethyl ester for subtilisin A-catalysed peptide synthesis. *Org. Biomol. Chem.* **10**, 6767–75 (2012).

09 — F Clapés, P., Torres, J. L. & Adlercreutz, P. Enzymatic peptide synthesis in low water content systems: preparative enzymatic synthesis of [Leu]- and [Met]-enkephalin derivatives. *Bioorg. Med. Chem.* **3**, 245–55 (1995).

10 — Klibanov, A. M. Why are enzymes less active in organic solvents than in water? *Trends Biotechnol.* **15**, 97–101 (1997).

11 — Li, Z., Roccatano, D., Lorenz, M. & Schwaneberg, U. Directed evolution of subtilisin E into a highly active and guanidinium chloride- and sodium dodecylsulfate-tolerant protease. *Chembiochem* **13**, 691–9 (2012).

12 — Foophow, T., Tanaka, S., Koga, Y., Takano, K. & Kanaya, S. Subtilisin-like serine protease from hyperthermophilic archaeon *Thermococcus kodakaraensis* with N- and C-terminal propeptides. *Protein Eng. Des. Sel.* **23**, 347–55 (2010).

13 — Olędzka, G., Dąbrowski, S. & Kur, J. High-level expression, secretion, and purification of the thermostable aqualysin I from *Thermus aquaticus* YT-1 in *Pichia pastoris*. *Protein Expr. Purif.* **29**, 223–229 (2003).

14 — Ma, J.-N., Currier, E. A., Essex, A., Feddock, M., Spalding, T. A., Nash, N. R., Brann, M. R. & Burstein, E. S. Discovery of novel peptide/receptor interactions: identification of PHM-27 as a potent agonist of the human calcitonin receptor. *Biochem. Pharmacol.* **67**, 1279–84 (2004).

15 — Nishino, N., Xu, M., Mihara, H. & Fujimoto, T. Use of hexafluoroisopropyl alcohol in tryptic condensation for partially protected precursor of α-melanocyte stimulating hormone. *Tetrahedron Lett.* **33**, 3137–3140 (1992).

16 — Abrahmsén, L., Tom, J., Burnier, J., Butcher, K. A., Kssiakoff, A. & Wells, J. A.. Engineering subtilisin and its substrates for efficient ligation of peptide bonds in aqueous solution. *Biochemistry* **30**, 4151–9 (1991).

17 — Wu, Z. & Hilvert, D. Conversion of a protease into an acyl transferase: Selenolsubtilisin. *J. Am. Chem. Soc.* 4513–4514 (1989).

18 — Qin, X., Xie, W., Tian, S., Cai, J., Yuan, H., Yu, Z., Butterfoss, G. L., Khuong, A. C. & Gross, R. A. Enzyme-triggered hydrogelation via self-assembly of alternating peptides. *Chem. Commun. (Camb).* **49**, 4839–41 (2013).

19 — Kaspar, A. a & Reichert, J. M. Future directions for peptide therapeutics development. *Drug Discov. Today* **18**, 807–17 (2013).

20 — Ritzefeld, M. Sortagging: a robust and efficient chemoenzymatic ligation strategy. *Chemistry* 20, 8516–29 (2014).

21 — Nguyen, G. K. T., Wang, S., Qiu, Y., Hemu, X., Lian, Y. & Tam, J. P. Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis. *Nat. Chem. Biol.* **10**, 732–8 (2014).

22 — Liebscher, S., Kornberger, P., Fink, G., Trost-Gross, E.-M., Höss, E., Skerra, A. & Bordusa, F. Derivatization of antibody fab segments: a designer enzyme for native protein modification. *Chembiochem* **15**, 1096–100 (2014).




## Nederlandse Samenvatting

Het gebruik van peptiden en eiwitten als medicijnen, voedingsadditieven en ingrediënten voor cosmetica wordt steeds prominenter. De markt voor therapeutische peptiden groeit constant en de verwachting is dat deze groter dan $25 miljard zal zijn binnen twee jaar. Het aantal therapeutische peptiden in ontwikkeling groeit, evenals hun lengte, complexiteit en de vraag naar grotere hoeveelheden.

Een algemene technologie voor de productie van peptiden bestaat niet. Hedendaagse methoden voor peptideproductie zijn chemische synthese, productie in recombinante systemen en chemo-enzymatische synthese. De geprefereerde synthese route hangt af van de sequentie, lengte en eigenschappen van de peptide. Hoofdstuk 1 geeft een kort overzicht van deze methoden.

Chemo-enzymatische peptide synthese (CEPS) is een alternatieve methode voor de productie van middelgrote tot grote peptiden. Enzymatische koppeling van chemisch gesynthetiseerde peptidesegmenten combineert de voordelen van vastefasesynthese (bijvoorbeeld het inbouwen van niet-natuurlijke aminozuren) en enzymatische ligatie (geen racemisatie, geen beschermende groepen en het gebruik van een vloeibare fase met beter oplossend vermogen voor peptiden). Enzymatische ligaties in CEPS zijn regio- en stereospecifiek, waardoor het gebruik van meervoudige HPLC zuiveringsstappen en daarmee ook de kosten worden gereduceerd.

In het vorig decennium was de chemo-enzymatische peptide synthese gericht op het gebruik van hydrolytische enzymen in niet-waterige omstandigheden, met als doel het voorkomen van de meest dominante nevenreactie: hydrolyse van zowel producten als substraten. Toepassingen die gebruik maken van substraat of medium engineering waren erg in trek voor het synthetiseren van nieuwe enzymen die tolerant waren voor organische oplosmiddelen. Verscheidene subtilisine-achtige enzymen (subtilases) zijn geselecteerd in dit onderzoek (hoofdstuk 3) en zijn getest in hydrolyse- en synthesereacties onder een variatie van omstandigheden. Thermostabiele subtilases, verkregen door genome mining van extremofielen, blijken robuuste enzymen te zijn die in een mesofiele gastheer geproduceerd kunnen worden en daarna ook makkelijk kunnen worden gezuiverd met een enkele hitte behandelingsstap. Dergelijke enzymen kunnen dan worden gebruikt als hydrolases onder extreme reactie omstandigheden. Ze kunnen ook worden geprecipiteerd door een organisch oplosmiddel en gebruikt worden als poederpreparaten voor synthese van korte peptiden in mengsels van

watervrije organische oplosmiddelen (hoofdstukken 4 en 5). Van bijzonder groot belang is de nieuwe subtilase die door de grandioze DMF en DMSO tolerantie thermodynamisch gecontroleerde peptidesynthese mogelijk maakt, zoals beschreven in hoofdstuk 6. Ondanks de recente ontwikkelingen in watervrije enzymologie zou de enzymatische peptidesynthese in waterige oplossingen vele voordelen hebben. De verhoogde oplosbaarheid van peptiden in gebufferde oplossingen en de afwezigheid van beschermende groepen voor de zijketens zou de koppeling van zeer lange peptiden mogelijk maken. Thiol-subtilisines en selenol-subtilisines zijn ontworpen als ligases voor chemo-enzymatische peptidesynthese in waterige oplossingen.

Van alle nieuwe biokatalysatoren die beschreven worden in dit proefschrift, heeft peptiligase, een sterk geëngineerde thiol-subtilisine, de potentie om een chemo-enzymatische peptidesynthese technologie te bewerkstelligen als geprefereerde methode voor synthese van (farmaceutische) peptiden en eiwitten op een kostenefficiënte en milieuvriendelijke manier (hoofdstuk 7). Wij hebben hoge koppelingsefficiënties (hoge S/H ratio's) verkregen voor peptiligase gekatalyseerde ligaties, met name voor lange, onbeschermde peptiden in gebufferde oplossingen. Deze kinetisch gecontroleerde reacties, die een geactiveerde acyl-donor en een geschikte nucleofiele acyl-acceptor vereisen, zijn snel en hebben maar een kleine hoeveelheid enzym nodig. De acyl-donor is over het algemeen N-terminaal beschermd en C-terminaal geactiveerd met een Cam-Leu ester. Een amine wordt gebruikt als nucleofiel. Peptiligase heeft zes substraatbindingsholtes die een rol spelen in substraatherkenning en katalyse. De volgende stap zou het onderzoek naar de substraatrange van peptiligase zijn en het verbreden van deze range met gerichte evolutie (directed evolution) en computational design. Het is zeer onwaarschijnlijk dat een enkele peptiligase-variant genoeg zal zijn om alle mogelijke combinaties van peptidesubstraten te kunnen koppelen, dus de engineering zal zich focussen op het construeren van een ligase-toolbox die in staat is om alle mogelijke, praktisch relevante sequenties te kunnen koppelen. Deze enzymen zouden als "restrictie ligases" beschreven kunnen worden, ontworpen voor een specifieke combinatie van acyl-donor en nucleofiel. Met deze enzymen zouden one-pot-reacties voor peptide-assemblage mogelijk worden gemaakt die, naast toepassing in de farmaceutische industrie, ook gebruikt kunnen worden voor nieuwe materialen, zoals zelf-assemblerende biopolymeren.

Naast synthese van bekende peptidesequenties ontstaat er ook een steeds grotere interesse in modificatie reacties die leiden tot een grotere variatie in structuur en biologisch gedrag. Bio-activiteit en farmacokinetiek kunnen worden





gemodificeerd door het maken van peptide conjugaties en peptiden die niet natuurlijke aminozuren bevatten of door de ontwikkeling van cyclische peptiden. Zoals is laten zien in hoofdstuk 7, is peptiligase in staat om een lineaire peptide van 21 aminozuren kop-staart te cycliseren als enige product, gebruikmakende van een hoge peptideconcentratie en korte reactietijden. Andere interessante toepassingen zullen opkomen als de substraatspecificiteit van peptiligase kan worden uitgebreid en zo een grote variëteit van peptiden met verschillende herkenningssequenties gekoppeld kunnen worden door één enzym. Dit maakt "traceless" modificatie mogelijk: peptiden modificeren en labelen zonder dat een herkenningssequentie nodig is. Deze techniek is nog niet mogelijk, omdat de huidige ligases en transpeptidases een herkenningssequentie motief nodig hebben (sortase, butelase, trypsiligase). Selectieve peptide ligases zullen ook de onbeschermde koppeling van peptide segmenten mogelijk maken die geproduceerd zijn door fermentatie. Dit opent de deur naar eiwit en antilichaam-medicijn-conjugaten.

Concluderend, blijkt dat de protein engineering van subtilisine en thiol-subtilisine varianten veel potentie heeft, zelfs nu er al 25 jaar verstreken is sinds de ontdekking van het eerste enzym subtiligase. De peptide ligase (peptiligase) die in dit proefschrift wordt beschreven, is een robuust, thermostabiel enzym met een hoge synthese/hydrolyse ratio en kan de sleutel tot succesvolle industriële chemo-enzymatische peptidesynthese technologie zijn. Verdere engineeringstudies, in combinatie met de voortschrijdende in silico designtechnieken, zouden het op maat maken van biokatalysatoren voor specifieke reactievereisten mogelijk moeten maken.

# Dankwoord / Acknowledgements

My journey to Groningen started with an interest in enzymes. This idea has been crystallizing for a while in Zagreb, and talking to you Ivaylo and Maja has set the process in motion. Hvala!

First of all I would like to thank my promotor for accepting me as a team member. Dear prof. Janssen, I cannot express enough my gratitude to you for all your support during all these years. I have learned a lot and my respect for you as a scientist and as a person has grown even more. Thank you for all that you have thought me! I am proud to call you my promotor. I truly hope our scientific collaboration will continue in the future.

I would also like to thank prof. Marco Fraaije, for all the support and for recognizing a potential student assistant in me.

I owe my gratitude to the members of the reading committee: Prof. dr. Dirk Jan Slootboom, Prof. dr. Rob M. J. Liskamp and Prof. dr. Oscar P. Kuipers for reviewing this thesis. Few more people were involved on the project from the very start. Dear Timo, we started collaboration during IBOS-II project (Integration of Biosynthesis and Organic Synthesis) eight years ago, working on enzymes for peptide synthesis. You did outstanding work (in Geleen) on enzymatic peptide synthesis in organic solvent and we (in Groningen) were looking for new enzymes that could do synthesis in organic solvents and in water. We were efficient team (even on those distances) and since my move to the South and to the lab one corridor away we became even more efficient, now both as EnzyPep labteam (together with Michel, Mathijs and Marcel). I am looking forward to our new discoveries and possibilities of CEPS technology, to the new better enzymes and to the delicious tips from your previous career (as a cook) and many new tips from your other careers in the future!

Dear PQ, (P.J.L.M. Quaedflieg), our meetings, especially last years in Groningen were a great pleasure not only because of scientific progress, but because they felt as reunion of friends. I always appreciated your healthy appetite, especially for crème brûlée. It took your skills to bring EnzyPep to its feet. Thank you for adding another Slavic language to your polyglot list!

I would also like to acknowledge the former IBOS-II members with whom I spend many hours in fruitful scientific discussions during the project meetings: Jan-Metke van der Laan (DSM Delft), from Wageningen University dr. Petra Vossenberg, Prof. dr. ir. Hans Tramper and dr. Rik Beeftink and from Radboud University in Nijmegen dr. Roseri de Beer-Roefles and Prof. Floris P.J.T. Rutjes.

Dear Wu, thank you for being my PostDoc. This was not always an easy project and you had a difficult task to guide me through sometimes frustrating situations, especially with translating the scientific experiments into a publication. I am proud that you achieved your dream of academic career in China.

Next, I would like to thank my dear paranymphs.

Dear Gosia, I am happy to have you as a friend. You stood by me during my difficult times and also during my most happy moments. Especially when I did not use the words to ask you to... Because you knew what was best for me. Thank you for that. We share a great friendship, and moving away from Groningen made me realize just how precious that friendship is. We shared so many moments in now historical Nijenborgh 4... there is not enough paper to name them all, but they are all carefully stored in my heart. Thank you for being my communication specialist. With your sharp analysis you kept my often wondering thoughts in order. I imagine that you often felt as if reading a book of James Joyce. So thank you for the patience. I wish you all the best, wherever the future brings you. And I will keep my phone with me.

Dear Nina, I still remember your first visit to Groningen. I am glad you chose Groningen and that you became my dear friend. Talking to you made me feel better every time. You are a great partner when things need to be organized and put in motion. Your compassion and determination makes things happen. I hope I will be able to give you back some of the support that I received during my finalizing year in Groningen. A big thank to you and Geoffrey for spoiling me as a guest... Tsjerk and I still talk about the brunch at your place...we call it a brunch with a big letter B now. Thank you for being my paranymph and thank you for being my friend.

Dear Anette, when I think of my first days in Groningen I think of you. You were there from day one with a smile and friendship to offer. We explored the city together. You showed me all the hidden shops and restaurants. You are curious and you truly enjoy life with full lungs and you seek adventures (or they find you)... so me, a less adventurous type experienced fantastic adventures together with you! The last one was in Leipzig! And I am looking for many more to come. Dear Anette, you helped me feel a part of the group and you held the group together with just being yourself.



Dankwoord / Acknowledgements

Dear Marcelo, the sixth chapter would not have been finished without your work. Thank you for that. I wish you all the best in the future!

Dear Hein, I appreciate all your help and our scientific and non-scientific talks. I believe we will still meet in the university corridors in Groningen.

Dear students! It was a great pleasure to watch your progress in the lab, and it made me even more proud that some of you continued your Bachelor, Master and even PhD in our group! Thank you Linda, Jeroen, Baastian, Michel, Nadia, Marcel and Jory. Special thank goes to Jeroen for his extra summer work on the project and for translation of summary to Dutch. I would not have done it without you. Good luck with your PhD! Jory, I am also very grateful for your work on the project as a part of your Bachelor research. Good luck with your Master!

Dear Piet and Chris, thank you for all your technical help during my PhD, for the discussions during the coffee break and for taking care of the fish. Dear Sandra, thank you for handling all my administrative issues.

Dear Marcus and Anett! Thank you for welcoming me in the group. From you I learned kindness and a lot of science. Dear Hanna, thank you for your support and great polish dinners and parties! Dear Danny, your positive energy is contagious! Thank you for introducing me to the borrel tradition! I steel keep it! Dear Hugo, work smarter and not harder was your motto, but unfortunately somehow “not” got lost in my translation. I am still looking for it. Dear Kuba, thank you for your brisk comments, all the delicious bbq’s and parties at your place. It was great to have you as an office neighbor.

Dear Arif, you were my longest officemate in Groningen. We shared office, project and many moments together. I hope you will find your way back to Groningen. Dear Roga, we shared many long days at the office and it was fantastic that we met last year again in Indonesia. I wish you all the luck with your job! Also, Alja and Maarten, thank you for your introduction to the Dutch culture! Dear Cynthia, Thai and Misun, you are the new generation of my officemates. Thank you for your smiles and stories from the happy office. Good luck in finishing your thesis.

Also, I want to thank all the past and current members of the group for all the discussions, support, advice and fun that we shared over the past years. Dear Willem, Edwin, Wiktor, Hemant, Patricia, Elisa Ana, Remko, Elvira, Laura, Sebastiaan,





Antonija, Dana, Jan, Marzena, Hesam, Nikola, Max, Hasan, Samin, Matt, Robert, Marleen, Aline, Yaser, Kamil, Alex, Ciprian, Thomas, Andy and Arne thank you! Dear Peter, I will miss our talks next to the Senseo machine.

Also I would like to acknowledge the friday borrel team of Nijenborgh 4. It was a great place to share ideas and beers.

Dear Jurica, this chapter would not be complete without you. Although we met our half-lifetime ago, it was in Groningen that we became close friends. Not because we were two fellow chemists seting for Groningen from the same country, but because we could share and laugh about our experiences of Groningen, and when this was not enough we would set to explore other countries and have a lot of food and fun. These travels are one of my dearest memories. If you wonder where we will meet again...it will for sure be either in a restaurant (food for you) or in a coffee place (coffee for me).

To my Ladies First, a big thank you for all the hours we spend on the water rowing, but also thank you for all the hours we spend together having fun. Thank you Meirav, Mariken, Wemke, Irene, Sietske, Marijke, Rianne, Anouk and thank you Wilfred for coaching me.

Draga  Jasmina, Rosana, Martina hvala na podršci!

In the end, I would like to express deep gratitude to my family.
My family in Croatia (Mama i tata! Puno hvala za sve!) and my frisian family in the Netherlands (Tankewol!). Special thanks to my brother: Ivo, I would not achieve so much without you.

Lieve Tsjerk,
we became a new family and our journey together just starts.
Thank you for being here for me.